



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY

INSTITUTE OF MATHEMATICS

**RESTAURACE POŠKOZENÝCH AUDIOSIGNÁLŮ POMOCÍ
ŘÍDKÝCH REPREZENTACÍ**

RESTORATION OF DEGRADED AUDIOSIGNALS USING SPARSE REPRESENTATIONS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Ondřej Mokrý

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Mgr. Pavel Rajmic, Ph.D.

BRNO 2017

Zadání bakalářské práce

Ústav: Ústav matematiky
Student: **Ondřej Mokrý**
Studijní program: Aplikované vědy v inženýrství
Studijní obor: Matematické inženýrství
Vedoucí práce: **doc. Mgr. Pavel Rajmic, Ph.D.**
Akademický rok: 2016/17

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Restaurace poškozených audiosignálů pomocí řídkých reprezentací

Stručná charakteristika problematiky úkolu:

Práce se týká problematiky restaurace audiosignálů, a to při znehodnocení výpadkem vzorků. Práce bude porovnávat starší metody pro restauraci signálů s novějšími, které jsou založeny na tzv. řídkosti signálů. V rámci práce budou rovněž navrženy modifikace některých metod za účelem zlepšení kvality rekonstrukce.

Cíle bakalářské práce:

Student prostuduje model znehodnocení audiosignálu při výpadku vzorků. Student analyzuje tento problém a seznámí se s obvyklými metodami pro jeho řešení, jejich vlastnostmi. Seznámí se rovněž s moderními metodami založenými na tzv. řídkosti signálů a navrhne související metodu restaurace a její modifikace. Metodu/metody poté implementuje v MATLABu a otestuje na simulovaných i reálných nahrávkách. Výstupem budou jednak výsledky objektivního srovnávání (SNR), jednak subjektivní hodnocení studenta.

Seznam doporučené literatury:

HRBÁČEK, R., RAJMIC, P., VESELÝ, V. a ŠPIŘÍK, J. Řídké reprezentace signálů: úvod do problematiky, Elektrevue, 2011. ISSN 1213-1539.

ŠPIŘÍK, J., RAJMIC, P. a VESELÝ, V. Reprezentace signálů: od bází k framům, Elektrevue, 2010. ISSN 1213-1539.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2016/17

V Brně, dne

L. S.

prof. RNDr. Josef Šlapal, CSc.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

Abstrakt

Bakalářská práce je zaměřena na problém doplnění úseku chybějících dat audiosignálu. Audiosignál je pomocí diskrétní Gaborovy transformace převeden na řídký vektor. Problém doplnění dat při zachování řídkosti reprezentace je formulován jako optimalizační úloha, která je následně řešena algoritmem Douglas-Rachford. Oproti stávajícím přístupům je navíc algoritmus doplněn o návrh metody pro kompenzaci poklesu energie v úseku doplněných vzorků výsledného signálu.

Summary

This bachelor thesis is focused on the problem of inpainting a segment of missing samples in an audiosignal. The signal is represented as sparse vector using discrete Gabor transform. The problem of inpainting missing samples while preserving the sparsity of the representation is formulated as an optimisation task, which is then solved using Douglas-Rachford algorithm. In contrast with the state-of-the-art approaches, the algorithm is extended by proposing method for compensating the energy decrease which occurs in the restored signal.

Klíčová slova

audiosignál, doplňování chybějících dat, Gaborova transformace, framy, řídké reprezentace, proximální operátory, algoritmus Douglas-Rachford, kompenzace poklesu energie

Keywords

audiosignal, inpainting, Gabor transform, frames, sparse representations, proximal operators, Douglas-Rachford algorithm, energy decrease compensation

MOKRÝ, O. *Restaurace poškozených audiosignálů pomocí řídkých reprezentací*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2017. 58 s. Vedoucí doc. Mgr. Pavel Rajmic, Ph.D.

Prohlašuji, že jsem bakalářskou práci *Restaurace poškozených audiosignálů pomocí řídkých reprezentací* zpracoval samostatně s využitím uvedených zdrojů pod vedením doc. Mgr. Pavla Rajmice, Ph.D.

Ondřej Mokrý

Děkuji doc. Mgr. Pavlu Rajmicovi, Ph.D. za odborné rady a přívětivý přístup při vedení této bakalářské práce. Také děkuji Dušanu Kovaříkovi za to, že inicioval, motivoval a umožnil vznik použitých nahrávek.

Ondřej Mokrý

Obsah

1	Úvod	2
2	Značení	3
2.1	Normy vektorů a řídkost	3
2.2	Moore-Penroseova pseudoinverze	3
2.3	Vlastnosti signálů	5
3	Řídké reprezentace signálů	6
3.1	Báze	6
3.2	Framy	7
3.2.1	Gaborovy framy	7
3.3	Syntéza a analýza	10
3.4	Řídké reprezentace	11
3.5	Algoritmy pro hledání řídkých řešení	12
3.5.1	Orthogonal Matching Pursuit	12
3.5.2	ℓ_1 relaxace	13
4	Doplňování chybějících úseků audiosignálu	15
4.1	Starší a alternativní metody	15
4.1.1	Metoda autoregrese	15
4.1.2	Metoda na bázi vlastního obsahu	15
4.1.3	Metoda podobnostních grafů	16
4.2	Metoda založená na řídkosti	17
4.2.1	Princip metody	17
4.2.2	Formulace problému	18
4.2.3	Proximální operátory	19
4.2.4	Návrh optimalizačního algoritmu	21
4.2.5	Maximální řešitelná délka díry	22
4.2.6	Kompenzace poklesu energie rekonstruovaného signálu	26
4.2.7	Návrh implementace algoritmu	29
5	Experimenty a vyhodnocení	30
5.1	Maximální řešitelná délka díry	30
5.2	Algoritmus pro doplňování chybějících úseků audiosignálu	32
5.2.1	Návrh experimentu	32
5.2.2	Výsledky	34
6	Závěr	43
7	Seznam použitých zdrojů	44
8	Seznam použitých zkratk	47
9	Seznam příloh	48

1. Úvod

Restaurace audiosignálů je aktuální problematikou. Při přenosu signálu může dojít k chvilkové poruše, čímž vznikne výpadek vzorků, nebo máme k dispozici signál s chybějícími či znehodnocenými úseky – typickým příkladem je záznam na gramofonové desce, která je poškozená a digitalizovaný signál tak obsahuje rušivé praskání.

Problematickou doplňování chybějících úseků do audiosignálu (v anglické literatuře se tento problém restaurace audiosignálu obvykle označuje jako *audio inpainting*) se zabývali autoři již v 80. letech, jak ukazují články [8] nebo [17]. V posledním desetiletí se začaly rozvíjet mimo jiné právě metody založené na řídkosti, stěžejní je například práce [2].

Motivace pro použití řídkých reprezentací je zřejmá z fyzikální podstaty audiosignálu, konkrétně z harmonické struktury hudebního tónu. Díky tomu lze audiosignál lokálně vyjádřit jako součet nemnoha harmonických funkcí (sinus, kosinus) o různé periodě a amplitudě, tedy s vhodnou množinou vektorů (ukáže se, že v našem případě se bude jednat o tzv. Gaborův frame) lze získat řídkou reprezentaci audiosignálu, čehož při řešení problému restaurace audiosignálu využijeme.

Cílem této práce je upravení vhodného optimalizačního algoritmu pro problém doplnění chybějícího úseku dat v audiosignálu, návrh možných zlepšení a implementace výsledného algoritmu v MATLABu, včetně otestování na reálných nahrávkách. Prioritou není rychlost algoritmu, ale co nejvěrnější rekonstrukce poškozeného signálu.

Tématu bakalářské práce se věnují kapitoly 2 až 5. Ve druhé kapitole jsou zavedeny vybrané matematické pojmy, které se úzce týkají řešené problematiky, a jsou zde odvozeny některé základní vlastnosti. Třetí kapitola se pak zabývá řídkými reprezentacemi. Je zde zaveden aparát bází a framů se zvláštním přihlédnutím ke Gaborovým framům, následně je zde rozšířena motivace pro využívání řídkých reprezentací audiosignálů a jsou zde uvedeny základní algoritmy pro jejich hledání. Čtvrtá kapitola se zabývá metodami doplňování chybějících úseků audiosignálu. Zmíněny jsou vybrané starší a alternativní metody a podrobně je rozebrána metoda založená na řídkosti. Kompletně je zde také navržen algoritmus pro řešení problému restaurace chybějících vzorků. Pátá kapitola se pak zabývá návrhem, realizací a vyhodnocením experimentu, který testuje algoritmus navržený ve čtvrté kapitole na reálných nahrávkách.

2. Značení

2.1. Normy vektorů a řídkost

Definice 1. *Nechť \mathbb{V} je vektorový prostor nad \mathbb{C} . Funkci $\rho : \mathbb{V} \rightarrow \mathbb{R}$, splňující*

1. $\rho(a \cdot v) = |a| \cdot \rho(v) \quad \forall a \in \mathbb{C}, \forall v \in \mathbb{V}$
2. $\rho(u + v) \leq \rho(u) + \rho(v) \quad \forall u, v \in \mathbb{V}$
3. $\rho(v) \geq 0 \quad \forall v \in \mathbb{V}$
4. $\rho(v) = 0 \Leftrightarrow v = 0$,

nazýváme normou na vektorovém prostoru \mathbb{V} . Místo $\rho(v)$ zpravidla používáme označení $\|v\|$.

Definice 2. *Nechť $p \geq 1$ je reálné číslo. ℓ_p -normou vektoru $x \in \mathbb{C}^N$ rozumíme číslo*

$$\|x\|_p = \left(\sum_{i=0}^{N-1} |x_i|^p \right)^{\frac{1}{p}}. \quad (2.1)$$

Speciálně ℓ_∞ -normou vektoru x rozumíme číslo

$$\|x\|_\infty = \max(|x_0|, \dots, |x_{N-1}|). \quad (2.2)$$

Definice 3. *Nosičem vektoru $x \in \mathbb{C}^N$ rozumíme množinu indexů všech nenulových složek vektoru x a značíme její $\text{supp}(x)$.*

Definice 4. *Řídkostí vektoru $x \in \mathbb{C}^N$ rozumíme číslo*

$$\|x\|_0 = |\text{supp}(x)|, \quad (2.3)$$

kde $|\cdot|$ značí mohutnost množiny. Dále vektor $x \in \mathbb{C}^N$ nazveme k -řídkým, jestliže

$$\|x\|_0 \leq k. \quad (2.4)$$

Zřejmě ℓ_2 -norma je klasická Eukleidovská norma a ℓ_1 -norma, kterou budeme často využívat, je součtem absolutních hodnot všech složek daného vektoru. Jako ℓ_0 -norma se někdy označuje řídkost vektoru, ačkoliv se dle definice o normu nejedná, neboť nesplňuje první axiom.

2.2. Moore-Penroseova pseudoinverze

V textu budeme dále pracovat s maticemi, které nejsou invertibilní. Pro výpočty proto potřebujeme k obecné matici najít takovou, která bude mít obdobné vlastnosti, jako má inverzní matice k regulární matici. Poslouží nám právě hojně používaná Moore-Penroseova pseudoinverze.

Definice 5. Pseudoinverzí matice A rozumíme takovou matici A^+ , která splňuje

1. $AA^+A = A$,
2. $A^+AA^+ = A^+$,
3. $(AA^+)^* = AA^+$,
4. $(A^+A)^* = A^+A$.

Poznámka. Horním indexem $*$ v předchozí definici značíme matici hermitovsky transponovanou.

Význam pseudoinverze pro řešení soustav lineárních rovnic s nekonečně mnoha řešeními uvádí následující tvrzení:

Tvrzení 1. *Nechť $Ay = b$ je nedourčený systém lineárních rovnic. Pak $z = A^+b$ je řešením tohoto systému a navíc platí*

$$\|x\|_2^2 \geq \|z\|_2^2 \quad \forall x : Ax = b. \quad (2.5)$$

Poznámka. Nedourčeným systémem lineárních rovnic rozumíme takový, že počet neznámých převyšuje počet rovnic. Současně předpokládáme, že soustava není neřešitelná – má tedy nekonečně mnoho řešení. Tvrzení pak interpretujeme tak, že $z = A^+b$ je mezi všemi řešeními systému $Ax = b$ to s nejnižší energií.

Důkaz. Nejprve ukážeme, že $z = A^+b$ je řešení. Z definice pseudoinverze platí $(AA^+A)y = b$, což díky asociativitě násobení matic odpovídá $AA^+(Ay) = b$ a tedy $AA^+b = b$. Odtud je zřejmé, že A^+b je řešením dané soustavy.

Označme nyní $B = A^+A$. Potom platí $Bz = (A^+A)(A^+b) = (A^+AA^+)b = A^+b = z$ a $B^* = B$ (přímo z definice pseudoinverze) a pro všechna řešení x dané soustavy

$$\begin{aligned} z^*(x - z) &= (Bz)^*(x - z) = \\ &= z^*B^*(x - z) = \\ &= z^*(Bx - Bz) = \\ &= z^*(A^+Ax - z) = \\ &= z^*(A^+b - z) = \\ &= 0. \end{aligned} \quad (2.6)$$

Protože u^*v představuje v prostoru \mathbb{C}^N skalární součin vektorů u a v , můžeme energii řešení x nyní vyjádřit jako

$$\begin{aligned} \|x\|_2^2 &= \|z + x - z\|_2^2 = \\ &= \|z\|_2^2 + \|x - z\|_2^2 + 2z^*(x - z) = \\ &= \|z\|_2^2 + \|x - z\|_2^2 \geq \|z\|_2^2 \end{aligned} \quad (2.7)$$

a tvrzení je dokázáno. □

Další důležitou vlastností pseudoinverze, kterou budeme dále využívat, je rovnost

$$A^+ = A^*(AA^+)^+, \quad (2.8)$$

která je dokázána například v článku [4].

2.3. Vlastnosti signálů

V dalších kapitolách se budeme zabývat signály, které chápeme jako posloupnosti komplexních čísel; konkrétně pak konečné signály jsou restrikcemi komplexních posloupností na konečnou délku. Proto také nebudeme nijak rozlišovat mezi signálem x délky N , vektorem $x \in \mathbb{C}^N$ a diskrétní funkcí $\{0, 1, \dots, N-1\} \rightarrow \mathbb{C}$. Nebude-li řečeno jinak, budeme složky vektoru značit dolním indexem a indexovat od 0, tedy $x \in \mathbb{C}^N$ budeme zapisovat jako $(x_0, x_1, \dots, x_{N-1})$.

Na tomto místě zavedeme několik pojmů, které se audiosignálů úzce týkají a na které se budeme později odkazovat.

Definice 6. *Mějme signál $x \in \mathbb{C}^N$. Energií signálu x rozumíme číslo*

$$\|x\|_2^2 = \sum_{i=0}^{N-1} |x_i|^2. \quad (2.9)$$

Pro objektivní porovnání dvou signálů pak budeme používat míru SNR (z anglického *signal-to-noise ratio*) [2], definovanou následovně:

Definice 7. *Mějme signály $x, y \in \mathbb{C}^N$, $x \neq y$. Veličinou SNR s jednotkou dB pak rozumíme hodnotu*

$$\text{SNR}(x, y) = 10 \cdot \log \frac{\|x\|_2^2}{\|x - y\|_2^2}. \quad (2.10)$$

3. Řídké reprezentace signálů

3.1. Báze

Báze vektorového prostoru je základním pojmem proto, abychom mohli hovořit o reprezentacích vektorů. V tomto textu se omezíme na vektorové prostory konečné dimenze, protože prostory nekonečné dimenze pro řešený problém nejsou relevantní.

Definice 8. Podmnožina B vektorového prostoru \mathbb{V} se nazývá báze vektorového prostoru \mathbb{V} , právě když platí

1. B generuje prostor \mathbb{V} ,
2. B je lineárně nezávislá množina.

Dimenzí vektorového prostoru \mathbb{V} rozumíme počet prvků jeho báze B .

Každý vektor prostoru \mathbb{V} tedy můžeme vyjádřit jako lineární kombinaci prvků báze B a díky požadavku lineární nezávislosti B je navíc toto vyjádření jednoznačné. Pokud vektory báze uspořádáme jako sloupce matice (označíme ji opět B a budeme ji nazývat *matice báze B*), můžeme potom každý vektor $x \in \mathbb{V}$ zapsat jako

$$x = Bc, \quad (3.1)$$

kde c je vektor koeficientů. O vektoru c potom hovoříme jako o *reprezentaci vektoru (souřadnicích) x v bázi B* .

Definice 9. Bázi B vektorového prostoru \mathbb{V} dimenze N nazveme ortonormální, jestliže její vektory jsou ortonormální, tedy

$$\langle b_i, b_j \rangle = \delta_{i,j} \quad \forall i, j = 1, 2, \dots, N, \quad (3.2)$$

kde

$$\delta_{i,j} = \begin{cases} 0, & \text{pro } i \neq j \\ 1, & \text{pro } i = j. \end{cases} \quad (3.3)$$

Ortonormální bázi je například tzv. kanonická báze Eukleidovského prostoru, jejíž matice je jednotková. Pro nás jsou však takové báze zásadní kvůli následující vlastnosti jejich (ortonormálních) matic:

Tvrzení 2. Nechť A je ortonormální matice řádu N . Pak platí

$$A^{-1} = A^*. \quad (3.4)$$

Poznámka. Definice ortonormální matice nebyla přímo uvedena, neboť je ekvivalentní definici ortonormální báze.

Důkaz. Platí $A^*A = I$, kde I je jednotková matice. Skutečně prvek na pozici i, j součinu můžeme vyjádřit $\sum_{k=1}^N a_{i,k}^* a_{k,j}$, což odpovídá skalárnímu součinu i -tého a j -tého sloupce matice A ; výsledkem maticového součinu je tedy matice jednotková.

Zbývá pro úplnost ukázat, že $AA^* = I$. Protože obecně $(AB)^* = B^*A^*$ a hermitovská transpozice jednotkové matice je stále jednotková matice, platí také $I = (A^*A)^* = (A^*)^*A^* = AA^*$, tedy $AA^* = I$. \square

Význam tohoto tvrzení je osvětlen v části 3.3.

3.2. Framy

Protože chceme hledat reprezentace signálů v nějakém smyslu optimální (v našem případě nejřidší), potřebujeme zavést obecnější pojem než báze, protože reprezentace vektorů pomocí bází jsou jednoznačné. Používaným zobecněním bází vektorových prostorů jsou právě framy.

Definice 10. [13] *Množina (nejvýše spočetná) vektorů $\{d_k\}_{k \in I}$ v prostoru \mathbb{V} v něm tvoří frame, pokud existují konstanty $0 < A \leq B < \infty$ takové, že platí*

$$A\|x\|^2 \leq \sum_{k \in I} |\langle x, d_k \rangle|^2 \leq B\|x\|^2, \quad \forall x \in \mathbb{V}. \quad (3.5)$$

Konstanty A a B potom nazýváme dolní a horní mez framu. Supremum ze všech dolních mezí nazýváme optimální dolní mez, infimum ze všech horních mezí nazýváme optimální horní mez. Vektory $\{d_k\}_{k \in I}$ nazýváme atomy framu.

V této definici není přímo řečeno, že množina $D = \{d_k\}_{k \in I}$ musí být generátorem prostoru \mathbb{V} , ovšem plyne to z ní. Pokud by totiž tato množina prostor \mathbb{V} negenerovala, existoval by jistě nenulový vektor $x \in \mathbb{V}$ kolmý k podprostoru generovanému množinou D , tudíž by platilo

$$\langle x, d_k \rangle = 0 \quad \forall k \in I. \quad (3.6)$$

Pro takový vektor x by ovšem pro $A > 0$ nemohla být splněna podmínka

$$A\|x\|^2 \leq \sum_{k \in I} |\langle x, d_k \rangle|^2 = 0. \quad (3.7)$$

Frame tedy vždy musí být generátorem prostoru \mathbb{V} .

Definice 11. *Frame nazveme těsným, pokud pro jeho optimální meze platí $A = B$. Pokud navíc $A = B = 1$, hovoříme o Parsevalově framu.*

Těsné framy jsou pro nás důležité svojí blízkostí ortogonálním bázím, což ukážeme v části 3.3.

3.2.1. Gaborovy framy

Gaborovy framy souvisí s diskretní Fourierovou transformací, protože atomy Gaborova framu jsou blízké bázovým vektorům, pomocí kterých je vyjádřena právě Fourierova transformace. Obdobně jako v článku [11] tedy nejdříve zavedeme právě diskretní Fourierovu transformaci.

Definice 12. *Nechť $x = (x_0, x_1, \dots, x_{N-1}) \in \mathbb{C}^N$. Diskretní Fourierovou transformací rozumíme zobrazení $\mathcal{F} : \mathbb{C}^N \rightarrow \mathbb{C}^N$ dané předpisem*

$$\mathcal{F}\{x\}_m = \hat{x}_m = \sum_{n=0}^{N-1} x_n e^{-2\pi i m n / N}, \quad m = 0, 1, \dots, N-1. \quad (3.8)$$

V konstrukci Gaborových framů budeme dále potřebovat operátory translace a modulace, definované následovně:

Definice 13. *Nechť $k, l \in \{0, 1, \dots, N-1\}$.*

Translací T_k vektoru $x \in \mathbb{C}^N$ rozumíme vektor

$$T_k(x) = (x_{n-k})_{n=0}^{N-1}. \quad (3.9)$$

Modulací M_l vektoru $x \in \mathbb{C}^N$ rozumíme vektor

$$M_l(x) = (x_n e^{2\pi i l n / N})_{n=0}^{N-1}. \quad (3.10)$$

Poznámka. Index $n-k$ v předpisu pro translaci vektoru x délky N chápeme modulárně, tedy jako $(n-k) \bmod N$.

Modulaci můžeme na základě diskrétní Fourierovy transformace považovat též za frekvenční posun, protože platí

$$\begin{aligned} \widehat{M_l(x)}_m &= \sum_{n=0}^{N-1} (x_n e^{2\pi i l n / N}) e^{-2\pi i m n / N} = \\ &= \sum_{n=0}^{N-1} x_n e^{-2\pi i (m-l) n / N} = \\ &= \widehat{x}_{m-l}. \end{aligned}$$

Skutečně se tedy jedná o posun ve spektru po aplikaci Fourierovy transformace.

Kombinací operátoru translace a operátoru modulace ve smyslu předchozí definice pak dostaneme operátor

$$\pi_{k,l} : \mathbb{C}^N \rightarrow \mathbb{C}^N, \quad \pi_{k,l}(x) = M_l(T_k(x)). \quad (3.11)$$

Použitím tohoto operátoru již můžeme konečně definovat Gaborův systém.

Definice 14. *Nechť $\varphi \in \mathbb{C}^N$, $\varphi \neq 0$ a $A \subseteq \{0, 1, \dots, N-1\} \times \{0, 1, \dots, N-1\}$. Množinu*

$$\{\pi_{k,l}(\varphi)\}_{k,l \in A} \quad (3.12)$$

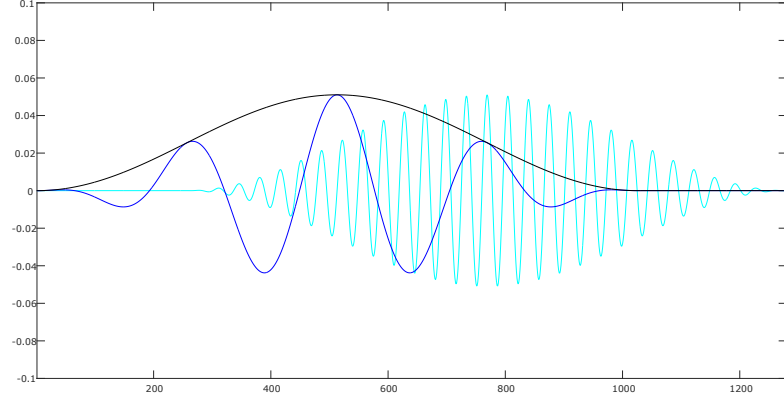
nazýváme Gaborův systém generovaný oknem φ . Pokud navíc tato množina generuje \mathbb{C}^N , nazýváme ji Gaborův frame.

Vektor φ , u kterého navíc předpokládáme, že jeho nosič je výrazně kratší než délka transformovaného signálu, pak nazýváme *okno Gaborova framu*. Gaborův frame se tedy skládá z translací a modulací okna φ , tzn. jeho časových a frekvenčních posunů. Hodnoty posunů udává volená množina A . Zde je vidět souvislost s Fourierovou transformací, která vyjadřuje vektor pomocí lineární kombinace funkcí sinus a kosinus o různých frekvencích, což v terminologii Gaborových framů odpovídá frekvenčním posunům konstantní funkce.

Příkladem funkce φ je v analýze audiosignálů často používané Hannovo okno [7], které je pro délku nosiče N definováno předpisem

$$w_n = \sin^2 \left(\frac{\pi n}{N-1} \right), \quad n = 0, 1, \dots, N-1. \quad (3.13)$$

Atomy Gaborova framu generovaného tímto oknem ilustruje obrázek 3.1.



Obrázek 3.1: Ukázka atomů Gaborova framu. Funkcí φ je zde Hannovo okno s délkou nosiče $w = 1024$ vzorků. Černě je zobrazen atom $\pi_{0,0}(\varphi)$ (tedy s nulovým posunem a nemodulovaný), modře atom $\pi_{0,4}(\varphi)$ (tedy s nulovým posunem, ovšem modulovaný s parametrem modulace $l = 4$), tyrkysově atom $\pi_{256,29}(\varphi)$ (tedy s posunem $k = 256$ vzorků a modulovaný s parametrem $l = 29$). Vykreslena je vždy reálná složka příslušného atomu.

Gaborova transformace

Vyjádření vektoru v reprezentaci Gaborova framu získáme diskrétní Gaborovou transformací, někdy zvanou též krátkodobá Fourierova transformace.

Definice 15. [16] Nechť $x = (x_0, x_1, \dots, x_{N-1}) \in \mathbb{C}^N$, $\varphi \in \mathbb{C}^N$, $a, M \in \mathbb{N}$ tak, že $L = \frac{N}{a}$ je celé číslo. Diskrétní Gaborovou transformací rozumíme zobrazení $\mathcal{G} : \mathbb{C}^N \rightarrow \mathbb{C}^{ML}$ dané předpisem

$$\mathcal{G}\{x\}_{m,l} = C_{m,l} = \sum_{n=0}^{N-1} x_n \pi_{al, -mN/M}(\overline{\varphi})_n, \quad (3.14)$$

kde $m = 0, 1, \dots, M-1$, $l = 0, 1, \dots, L-1$ a index $n - a \cdot l$ chápeme modulo N .

Poznámka. Index $n - a \cdot l$ je důsledkem translace n -té složky vektoru φ s parametrem $a \cdot l$, která je součástí použitého operátoru π . Kdybychom tento index neuvažovali modulo N , přesáhli bychom definiční obor okna φ , což není možné.

Vidíme tedy, že pro dané l a $M = N$ je Gaborova transformace vektoru x rovna Fourierově transformaci vektoru $\overline{T_{al}(\varphi)} \cdot x$, kde násobení rozumíme po složkách. Skutečně pro $m = 0, 1, \dots, M-1$ můžeme Gaborovu transformaci rozepsat jako

$$\mathcal{G}\{x\}_{m,l} = \sum_{n=0}^{N-1} x_n \pi_{al, -mN/M}(\overline{\varphi})_n = \sum_{n=0}^{N-1} x_n \overline{\varphi_{n-al}} e^{2\pi i(-mN/M)n/N} \quad (3.15)$$

a dosazením předpokladu $M = N$ a pevnou volbou l dostaneme

$$\mathcal{G}\{x\}_{m,l} = \sum_{n=0}^{N-1} x_n \overline{\varphi_{n-al}} e^{-2\pi i m n / N} = \mathcal{F}\{\overline{T_{al}(\varphi)} \cdot x\}_m. \quad (3.16)$$

Poznámka. Výsledek Gaborovy transformace zde chápeme jako jednorozměrný, neboť druhou souřadnici jsme předem pevně zvolili.

Inverzní Gaborova transformace

V aplikacích hledáme řídkou reprezentaci v Gaborově framu, ovšem jako výstup potřebujeme opět jednorozměrný signál. Pro matici koeficientů C rozměrů $M \times L$, okno φ a parametr a pak podle [16] definujeme inverzní diskrétní Gaborovu transformaci $\mathcal{G}^{-1} : \mathbb{C}^M \times \mathbb{C}^L \rightarrow \mathbb{C}^N$ předpisem

$$x_n = \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} C_{m,l} \pi_{al, mN/M}(\varphi)_n, \quad n = 0, 1, \dots, N-1. \quad (3.17)$$

Závěrem poznamenejme, že ačkoliv jsme koeficienty v Gaborově framu zavedli jako matici c rozměrů $M \times L$, v následujících kapitolách budeme hovořit o vektoru koeficientů c délky $M \cdot L$, který vznikne seřazením prvků matice C postupně po sloupcích.

3.3. Syntéza a analýza

V této části shrneme souvislosti mezi signály a jejich reprezentacemi. Vycházíme z vyjádření

$$x = Dc, \quad (3.18)$$

kde $x \in \mathbb{C}^N$ je signál délky N , D je matice rozměrů $N \times M$ a $c \in \mathbb{C}^M$ je vektor koeficientů. Matici D pak nazýváme *slovník* a rovnici (3.18) označujeme jako *syntézu signálu x z koeficientů c pomocí slovníku D* .

Pokud však známe signál a hledáme koeficienty reprezentace ve slovníku D , potřebujeme najít k rovnici (3.18) obdobné vyjádření

$$c = \hat{D}x, \quad (3.19)$$

které budeme nazývat *analýzou signálu*. Sloučením rovnic (3.18) a (3.19) tedy vidíme, že pro známou matici D hledáme matici \hat{D} splňující $\hat{D}D = I$, kde I je jednotková matice. Hledání takové matice rozdělíme na případy, kdy slovník D je báze, obecný frame a těsný frame.

Báze

Zřejmě jestliže D je bází prostoru \mathbb{C}^N , příslušná matice je čtvercová a regulární a tedy $\hat{D} = D^{-1}$. Pokud je navíc D ortonormální, pak $\hat{D} = D^*$, jak bylo dokázáno v tvrzení 2.

Framy

Protože framy nejsou na rozdíl od bází lineárně nezávislé množiny, nenajdeme reprezentaci c jednoznačně. Jako míra nejednoznačnosti nám slouží tzv. *redundance framu*, což je podíl $\frac{M}{N}$ počtu řádků a počtu sloupců matice D . Nyní ukážeme, že jako matici \hat{D} můžeme zvolit pseudoinverzi matice D . Vynásobením rovnice (3.18) zleva maticí DD^+ a využitím definice pseudoinverze dostaneme

$$DD^+x = DD^+Dc = Dc = x, \quad (3.20)$$

což je přímo požadovaná vlastnost.

V případě, kdy D je těsný frame s optimálními mezemi $A = B = K$, platí pro jeho pseudoinverzi [9]

$$D^+ = \frac{1}{K} D^*. \quad (3.21)$$

Zde právě vidíme souvislost těsných framů a ortonormálních bází – matici pro analýzu signálu dostaneme pouhou hermitovskou transpozicí matice pro syntézu a vynásobením vhodnou konstantou. Je-li navíc těsný frame Parsevalův, platí $A = B = 1$ a tedy $\hat{D} = D^+ = D^*$, což je stejný výsledek jako u ortonormálních bází.

Gaborovy framy

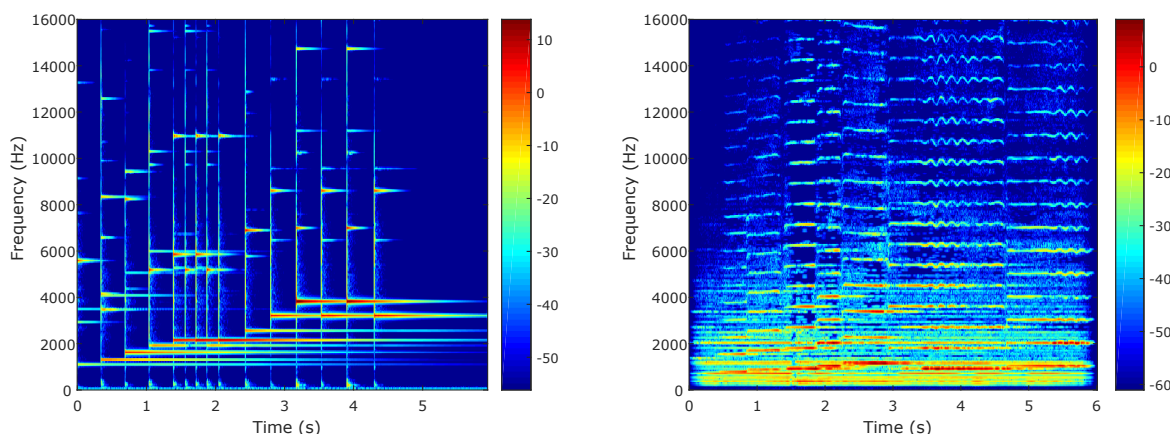
Analýzu signálu v případě, kdy slovník D je Gaborův frame, neprovádíme v řešeném problému pomocí násobení vhodnou maticí, ale vektor c určujeme po složkách Gaborovou transformací, definovanou vzorcem (3.14). Syntéza signálu z koeficientů pak odpovídá inverzní Gaborově transformaci dle vzorce (3.17).

3.4. Řídké reprezentace

V úvodní kapitole jsme zavedli řídkost vektoru a co znamená, když je vektor k -řídký (definice 4). Nyní situaci zjednodušíme a budeme vektor $x \in \mathbb{C}^N$ nazývat řídkým, jestliže splňuje $\|x\|_0 \ll N$, tzn. počet jeho nenulových složek je výrazně menší než N .

Stěžejním předpokladem v našem přístupu k restauraci audiosignálů je existence jejich řídkých reprezentací. Ukazuje se, že takové řídké reprezentace dostaneme právě využitím Gaborových framů. Snadno lze tuto skutečnost ozřejmit na hudebních signálech. Hudební zvuk – tón – je totiž z fyzikálního pohledu součtem nemnoha harmonických frekvencí, po Gaborově transformaci tedy musí být vektor koeficientů řídký.

Strukturu hudebního signálu, ze které plyne řídkost reprezentace v Gaborově framu, ilustrují následující spektrogramy, které v decibelech zobrazují zastoupení jednotlivých frekvencí (svislá osa) v signálu v daném čase (vodorovná osa).



Obrázek 3.2: Vlevo spektrogram signálu Glockenspiel (zvonkohra). Vpravo spektrogram záznamu houslí a violy.

Přestože v analyzovaném záznamu houslí a violy je zastoupeno výrazně více frekvencí, je na něm zřetelná řídká struktura harmonických frekvencí, ze kterých jsou jednot-

livé tóny. Protože algoritmus restaurace bude založen právě na řídkosti, předpokládáme, že pro signály s řidší reprezentací dosáhneme lepších výsledků.

3.5. Algoritmy pro hledání řídkých řešení

Ukázali jsme, že Gaborovou transformací dostaneme řídkou reprezentaci audiosignálu. Také ovšem víme, že vyjádření pomocí framů není (na rozdíl od bází) jednoznačné. Protože hledáme reprezentaci nejřidší, můžeme hovořit o řešení optimalizační úlohy

$$\arg \min_x \|x\|_0 \text{ vzhledem k } Dx = y, \quad (3.22)$$

kde $Dx = y$ je v tuto chvíli obecná lineární podmínka.

Protože přesné řešení této úlohy nelze v přijatelně krátkém čase najít, následuje ukázka metod, které řešení aproximují s dostatečnou přesností.

3.5.1. Orthogonal Matching Pursuit

Algoritmus Orthogonal Matching Pursuit (OMP) slouží k aproximaci nejřidšího řešení soustavy lineárních rovnic $Dx = y$, tedy řešíme problém (3.22). OMP algoritmus pro řešení tohoto problému má dle [1] následující tvar:

Algoritmus 1: OMP algoritmus

```

1 Vstup:  $y, D$ 
2 Počáteční nastavení:
3 čítač iterací  $n = 0$ 
4 nosič řešení  $\Omega_0 = \emptyset$ 
5 residuum  $r_0 = y$ 
6 for  $n = 1, 2, \dots$  do
7    $c_n = D^T r_{n-1}$ 
8    $i = \arg \max_i |c_n^i|$ 
9    $\Omega_n = \Omega_{n-1} \cup i$ 
10   $x_n = \arg \min_u \|y - D_{\Omega_n} u\|_2$ , kde  $D_{\Omega_n}$  je matice obsahující ty sloupce matice  $D$ ,
    jejichž index je prvkem  $\Omega_n$ 
11   $r_n = y - D_{\Omega_n} x_n$ 
12 end
13 Výstup:  $x_n$  z poslední iterace
```

V prvním kroku každé iterace tedy dostaneme vektor c_n s informací o odchylkách sloupců matice D a residua z minulé iterace r_{n-1} . Ve druhém kroku najdeme index i maximálního prvku vektoru c_n , což znamená, že i -tý sloupec matice D má ze všech sloupců největší odchylku od residua r_{n-1} . Následně nosič hledaného řešení rozšíříme o index i a najdeme vektor s tímto nosičem, který je (ve smyslu Eukleidovské normy) nejméně vzdálen podmínce řešeného problému. V posledním kroku aktualizujeme residuum a pokračujeme další iterací.

Protože nosič vektoru x postupně navyšujeme, dostáváme v k -té iteraci k -řídský vektor x_k aproximující řešení soustavy $y = Dx$. Navíc v každé iteraci minimalizujeme největší složku rezidua r_n .

Algoritmus končí, pokud norma aktuálního residua je menší než předem zvolená mez, nebo pokud dosáhneme maximálního počtu iterací. U druhé podmínky má smysl pouze nižší hranice, než je počet sloupců matice D , v opačném případě bychom dostali $D_{\Omega_n} = D$ a další iterace by již nemohly řešení nijak ovlivnit.

3.5.2. ℓ_1 relaxace

Protože ℓ_0 -norma není konvexní funkcí, nelze pro úlohy, ve kterých tuto normu minimalizujeme, použít osvědčené metody konvexní optimalizace. Lze však ukázat, že za určitých podmínek jsou dostatečně blízké výsledky minimalizace ℓ_0 -normy a ℓ_1 -normy, která již konvexní funkcí je. Právě řešení úloh s ℓ_1 -normou místo původní řidkosti nazýváme ℓ_1 relaxací.

Podmínky ekvivalence

Je stále otevřeným tématem, kdy má relaxovaná úloha totožné řešení s úlohou původní. Nejčastěji se hledání omezuje na situace, kdy hledáme v nějakém smyslu optimální řešení soustavy lineárních rovnic. Podmínkou ekvivalence relaxované a původní úlohy je pak určitá vlastnost matice soustavy. Jako příklad nám poslouží následující tvrzení.

Definice 16. [13] *Vzájemná koherence matice A je definována jako největší normalizovaný skalární součin dvou různých sloupců matice A ,*

$$\mu(A) = \max_{1 \leq i, j \leq N, i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}, \quad (3.23)$$

kde a_i označuje i -tý sloupec matice A a N je počet jejích sloupců.

Tvrzení 3. [13] *Pokud má soustava $Ax = b$ řešení x splňující*

$$\|x\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu(A)} \right), \quad (3.24)$$

pak x je nutně nejřidší možné a je jediné takové. Navíc toto řešení lze dosáhnout ℓ_1 minimalizací.

V řešeném problému je maticí soustavy matice obsahující jako sloupce atomy zvoleného framu, pravá strana soustavy je původní signál a hledáme (nejřidší) vektor koeficientů v reprezentaci daného framu. Předchozí tvrzení nám říká, že pokud jsou atomy framu dostatečně nekoherentní a hledané řešení je dostatečně řídké, bude totožné s řešením relaxované úlohy. V praxi proto volíme framy, které nejsou příliš redundantní, protože v opačném případě bude koherence atomů vždy vysoká a kritérium tak nebude splněno.

Algoritmus Douglas-Rachford

Douglas-Rachfordův (DR) algoritmus slouží k nalezení řešení problému

$$\arg \min_x f(x) + g(x). \quad (3.25)$$

Podle [6] je obecný tvar DR algoritmu pro vyřešení takto formulovaného problému následující:

Algoritmus 2: Obecný tvar DR algoritmu

```
1 zvolíme  $\varepsilon \in (0, 1)$ ,  $\tau > 0$ ,  $y_0 \in \mathbb{R}^N$ 
2 for  $n = 0, 1, 2, \dots$  do
3    $x_n = \text{prox}_{\tau g} y_n$ 
4   zvolíme  $\lambda_n \in \langle \varepsilon, 2 - \varepsilon \rangle$ 
5    $y_{n+1} = y_n + \lambda_n (\text{prox}_{\tau f} (2x_n - y_n) - x_n)$ 
6 end
```

Zde ε , τ a λ jsou volené parametry, y_0 je počáteční iterace. Za podmínek

$$\text{dom}(f) \cap \text{dom}(g) \neq \emptyset \quad \wedge \quad \lim_{\|x\| \rightarrow \infty} f(x) + g(x) = \infty \quad (3.26)$$

posloupnost $\{x_n\}_{n=0}^{\infty}$ konverguje k hledanému vektoru, který minimalizuje funkci $(f + g)(x)$. Volbou parametrů pak můžeme ovlivnit rychlost konvergence algoritmu. V části 4.2.6 navíc ukážeme, že pro řešený problém má smysl uvažovat i DR algoritmus v modifikované podobě, kdy $\tau \in \mathbb{R}^N$ je vektor. Algoritmus poté konverguje k jinému bodu, než je-li parametr τ konstantní, což je níže řádně okomentováno.

DR algoritmus zdánlivě s ℓ_1 relaxací přímo nesouvisí, ovšem pro minimalizaci ℓ_0 -normy nemůže být použit, protože pro ℓ_0 -normu neznáme proximální operátor (tento termín je řádně zaveden v následující kapitole). Proto tento algoritmus použijeme právě pro relaxovanou úlohu s ℓ_1 -normou.

4. Doplnování chybějících úseků audiosignálu

4.1. Starší a alternativní metody

4.1.1. Metoda autoregrese

Metoda autoregrese, která je založená na predikci vzorků signálu pomocí vzorků předcházejících, patří k oblasti doplňování chybějících úseků audiosignálu k nejstarším – byla představena v 80. letech například v článcích [8] a [17]. V základní podobě této metody nejdříve vzorky ve známém úseku, který předchází díře v signálu, vyjádříme jako lineární kombinace prvků předcházejících (počet p prvků braných v potaz nazýváme řádem metody). Pro celý známý úsek se pak snažíme najít koeficienty lineární kombinace takové, že pokud je použijeme postupně pro vyjádření všech známých vzorků, bude součet kvadrátů odchylek těchto odhadů od skutečných vzorků minimální. Tyto koeficienty poté použijeme k odhadu chybějících prvků.

Lepšího výsledku dosáhneme, pokud současně provádíme odhad vzorků z levého a z pravého okolí díry a výsledné rekonstruované vzorky jsou vždy vhodnou kombinací těchto dvou odhadů (základní ukázkou takové vhodné kombinace může být vážený průměr dle pozice rekonstruovaného vzorku vzhledem k počátečnímu a koncovému prvku díry).

Pokročilejší verzi metody je adaptivní interpolace. Zde odhad koeficientů (tzv. autoregresní funkce) i samotných vzorků probíhá iteračně v rámci minimalizace jedné funkce. Nejdříve volíme počáteční vektor hledaných vzorků (například jako nulový vektor), pak pro celý úsek signálu najdeme autoregresní koeficienty minimalizující odchylku od známých vzorků, pomocí těchto koeficientů určíme nový vektor hledaných vzorků a proces opakujeme.

Výhodou metody autoregrese je její rychlost, lze ji použít i pro rekonstrukci v reálném čase. V případě použití iterační metody jsou obdržené výsledky oproti metodě základní lepší již při 3 iteracích, metoda tedy není výrazně pomalejší. Nevýhodou je, že pro delší úseky chybějících vzorků nejsou výsledky rekonstrukce uspokojivé – autoři na experimentech ukázali dobré výsledky u hudebního signálu s vzorkovací frekvencí 44 100 Hz a délkou díry 16 vzorků (což odpovídá 0,36 ms), u záznamu řeči s vzorkovací frekvencí 8 000 Hz pak dosáhli uspokojivých výsledků při doplnění 100 vzorků (12,5 ms).

4.1.2. Metoda na bázi vlastního obsahu

Metoda prezentovaná v článku [3] je, obdobně jako metoda autoregrese, založena na predikci – ze známého úseku signálu sbíráme informaci o pravděpodobném dalším průběhu, pomocí této informace poté rekonstruujeme chybějící vzorky. Ačkoliv metoda je primárně určena pro rekonstrukci řeči, na obdobném principu staví další moderní postupy rekonstrukce hudebního signálu; základní princip této metody zde proto uvedeme.

Základem je rozdělení signálu na překrývající se segmenty, které analyzujeme s využitím skrytého Markovova modelu – jedná se o analýzu používanou například při rozpoznávání řeči, kdy ze signálu extrahujeme významné charakteristiky (využití modelu je přiblíženo například v [19]). Touto analýzou získáme při postupném zpracovávání signálu

pro každý segment reprezentaci pomocí číselného vektoru a tyto reprezentace zaznamenáváme. Jakmile narazíme na segment obsahující chybějící vzorky, použijeme místo něj segment z dosud zaznamenaných, který splňuje dvě podmínky:

- odchylka nahrazovacího segmentu od segmentu původního v jeho nepoškozené části je minimální,
- pravděpodobnost posloupnosti segmentů v okolí díry včetně segmentu nahrazujícího je maximální.

Druhou podmínku ověřujeme právě na základě pozorování dostatečně dlouhého úseku před dírou v signálu. Využijeme zde předpokladu, že se v signálu jisté prvky opakují – při rekonstrukci záznamu řeči to mohou být opakující se slova či hlásky, v hudebním signálu by se pak jednalo například o skupiny tónů.

Vzhledem k tomu, že navržený algoritmus je k více vlastnostem řeči necitlivý, je po nalezení vhodného nahrazujícího segmentu ještě zařazeno několik operací:

- Nahrazujeme-li díru (v mluvené řeči) v samohlásce, je potřeba brát v úvahu její výšku jakožto výšku tónu. V takovém případě tudíž proběhne ještě modulace frekvence tak, aby výška odpovídala ostatním výskytům této samohlásky.
- Obdobně je upravena síla zvuku.
- Výsledný signál je také vyhlazen tím, že přechod mezi segmenty nepoškozenými a segmentem nahrazujícím díru v jejich překrývající se části je plynulý.

V provedeném experimentu byl metodou na bázi vlastního obsahu rekonstruován záznam řeči s vzorkovací frekvencí 8 000 Hz a úseky chybějících vzorků až do délky 240 ms (což odpovídá 1 920 vzorkům). Dle subjektivního hodnocení respondentů průzkumu byly poruchy v původním signálu vnímány jako *obtěžující* až *velmi obtěžující*, místa poruch po rekonstrukci signálu pak byla hodnocena jako *mírně obtěžující* až *rozpoznatelná, ale neobtěžující* [3, část Diskuze výsledků]. Na rozdíl od metody autoregrese však tento přístup neumožňuje rekonstrukci poškozeného audiosignálu v reálném čase.

4.1.3. Metoda podobnostních grafů

Tato moderní metoda, prezentovaná v [10], je založena na principu metody na bázi vlastního obsahu – signál rozložíme na segmenty a poté pro každý porušený segment hledáme dle určitých kritérií optimální neporušený, kterým jej nahradíme. Tento algoritmus však řeší úseky chybějících vzorků až do délky řádu sekund a je především určen pro rekonstrukci hudebního signálu. Předpokladem totiž je, že se dlouhé úseky v nevelkých obměnách opakují. V popové či rockové hudbě je tato podmínka splněna velice dobře a i v méně předvídatelných žánrech najdeme pro poškozený úsek dostatečně podobný nepoškozený.

Segmenty v této metodě jsou analyzovány pomocí Gaborovy transformace a dále pracujeme s jejich spektrogramy (struktura spektrogramu byla přiblížena v části 3.4). Právě podobnost jednotlivých spektrogramů je následně reprezentována grafem. Pomocí grafu pak algoritmus pro úsek s chybějícími vzorky hledá nepoškozený úsek signálu takový, že je maximální podobnost přechodových segmentů mezi porušenou a neporušenou částí signálu vzhledem k odpovídajícím přechodovým segmentům po aplikování náhradního úseku.

Specifika metody podobnostních grafů (PG) shrneme v následujícím srovnání s metodou na bázi vlastního obsahu (VO) [10, část Související práce]:

- Metoda VO, ač testována i pro delší úseky chybějících dat, je navržena pro signály s dírami do délky 40 ms (což je obvyklá délka v důsledku chybějících paketů dat při přenosu signálu). Metoda PG cílí na rekonstrukci děr délky řádu desetin až jednotek sekund.
- Pro metodu PG není stěžejní rovnost původního signálu v okolí díry a odpovídajících úseků náhrady. Navíc plynulost přechodu mezi neporušenými úseky signálu a náhradou má vyšší prioritu než zachování přesné délky signálu.
- Cílem metody VO je co nejpřesnější náhrada chybějících dat, metoda PG se naopak zaměřuje na minimální rozpoznatelnost náhrady, minimalizace odchylky rekonstruovaného signálu a signálu před jeho poškozením není prioritou.

Poznámka. Zřejmě metody jsou zaměřeny na rekonstrukci odlišných typů signálu, srovnáváme tedy jen takové vlastnosti, které jsou pro všechny metody určené pro restauraci audiosignálu společné.

4.2. Metoda založená na řídkosti

Metodou doplňování chybějících úseků audiosignálu založenou na řídkých reprezentacích se již zabývalo více autorů – například v článku [2] autoři řeší problém v podobě, která bude představena v této kapitole. Optimalizační úlohu, na kterou problém převedeme, však řeší odlišným algoritmem, konkrétně již zmiňovaným OMP.

4.2.1. Princip metody

Základním principem je nalezení řídké reprezentace nepoškozené části signálu pomocí redukováného slovníku a následná syntéza z nalezeného řídkého vektoru koeficientů pomocí slovníku neredukovaného. Podrobněji lze postup popsat následovně:

Nechť $y \in \mathbb{C}^N$ je pozorovaný signál s dírou na pozicích $u, u + 1, \dots, v$. Tento signál lze reprezentovat pomocí slovníku (v tomto případě Gaborova framu) D , který chápeme jako matici komplexních čísel rozměrů $N \times M$, kde M je celkový počet atomů framu.

1. Sestrojíme redukováný slovník D^r tak, že vypustíme řádky s indexy $u, u + 1, \dots, v$. Pomocí tohoto slovníku poté lze vyjádřit vektor y^r , který tvoří pouze neporušená část původního signálu y (je zřejmé, že počet atomů redukováného slovníku je M a délka atomů i signálu je $N - h$, kde h je počet chybějících vzorků v původním signálu).
2. Najdeme reprezentaci $x \in \mathbb{C}^M$ signálu y^r ve slovníku D^r , tedy vektor x splňující

$$D^r x = y^r \quad (4.1)$$

tak, aby byl nejřidší možný. Řídkost zde požadujeme na základě předpokladu, že reprezentace audiosignálu v Gaborově framu je obecně řídká.

3. Pomocí vektoru x a neredukovaného slovníku D syntetizujeme rekonstruovaný signál, což můžeme zapsat jako

$$\hat{y} = Dx. \quad (4.2)$$

Díky tomu, že vektor x splňuje rovnici (4.1), máme zaručeno, že příslušná část rekonstruovaného signálu \hat{y} se neliší od neporušené části signálu původního.

Poznámka. Signál y jsme zde uvažovali obecně komplexní, vzhledem k řešené problematice hudebních nahrávek se však v implementaci omezíme pouze na signály reálné.

Nyní postupně odvodíme podobu algoritmu, pomocí kterého realizujeme druhý krok uvedeného postupu.

4.2.2. Formulace problému

Problém hledání řídké reprezentace rekonstruovaného signálu lze formulovat jako optimalizační úlohu

$$\arg \min_x \|x\|_0 \text{ vzhledem k } D^r x = y^r, \quad (4.3)$$

y^r neporušená část původního signálu,

D^r část slovníku D odpovídající neporušené části signálu.

Uvažujme matici I^r , která vznikne z jednotkové matice vypuštěním řádků odpovídajících úseku chybějících vzorků v signálu. Potom

$$y^r = I^r y, \quad (4.4)$$

$$D^r = I^r D. \quad (4.5)$$

Abychom mohli použít Douglas-Rachfordův (DR) algoritmus, musíme problém formulovat ve tvaru (3.25). Tento tvar získáme, položíme-li

$$f(x) = \iota_{\{z: D^r z = y^r\}}(x), \quad (4.6)$$

$$g(x) = \|x\|_0. \quad (4.7)$$

Protože v DR algoritmu používáme proximální operátory funkcí $f(x)$ a $g(x)$ a pro ℓ_0 -normu proximální operátor neznáme, řešíme úlohu relaxovanou

$$\arg \min_x \|x\|_1 \text{ vzhledem k } D^r x = y^r. \quad (4.8)$$

Funkci $g(x)$ pak místo předpisu (4.7) definujeme jako

$$g(x) = \|x\|_1 = \sum_{i=0}^{N-1} |x_i|. \quad (4.9)$$

4.2.3. Proximální operátory

Proximální operátory jsou aparátem konvexní optimalizace, do které spadá řešení problému (3.25). V této části proximální operátor zavedeme a uvedeme či odvodíme konkrétní podoby, které budeme dále používat v návrhu algoritmu.

Definice 17. *Jestliže úloha*

$$\arg \min_{y \in \mathbb{R}^N} \frac{1}{2} \|x - y\|_2^2 + f(y) \quad (4.10)$$

má pro danou funkci f jednoznačné řešení, pak toto řešení značíme $\text{prox}_f x$ a operátor $\text{prox}_f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ nazýváme proximálním operátorem funkce f .

Proximální operátor tedy určuje takový vektor, který není příliš vzdálen (ve smyslu Eukleidovské normy) danému vektoru a současně minimalizuje danou funkci.

Tvrzení 4. [5] *Pro libovolné reálné číslo $\tau > 0$ je*

$$\text{prox}_{\tau \|\cdot\|_1}(x) = (\text{sgn}(x_i) \cdot \max\{|x_i| - \tau, 0\})_{i=0}^{N-1}. \quad (4.11)$$

Důkaz. Dosazením ℓ_1 -normy do definice proximálního operátoru dostaneme

$$\text{prox}_{\tau \|\cdot\|_1}(x) = \arg \min_{y \in \mathbb{R}^N} \frac{1}{2} \|x - y\|_2^2 + \tau \|y\|_1. \quad (4.12)$$

Objektivní funkci označíme $g(y)$ a vyjádříme

$$g(y) = \frac{1}{2} \sum_{i=0}^{N-1} (x_i - y_i)^2 + \tau \sum_{i=0}^{N-1} |y_i|. \quad (4.13)$$

Vzhledem k tomu, že funkce $g(y)$ je konvexní, pro nalezení minima je postačující splnění následujících podmínek:

Pro $y_i \neq 0$ je funkce $g(y)$ parciálně diferencovatelná podle y_i , pro minimum tedy platí

$$\frac{\partial g}{\partial y_i} = 0. \quad (4.14)$$

- Pro $y_i > 0$ je

$$\frac{\partial g}{\partial y_i} = (x_i - y_i) \cdot (-1) + \tau = 0 \quad (4.15)$$

a tedy

$$y_i = x_i - \tau. \quad (4.16)$$

Protože $y_i > 0$, platí předchozí rovnice pro $x_i > \tau$.

- Pro $y_i < 0$ je

$$\frac{\partial g}{\partial y_i} = (x_i - y_i) \cdot (-1) - \tau = 0 \quad (4.17)$$

a tedy

$$y_i = x_i + \tau. \quad (4.18)$$

Protože $y_i < 0$, platí předchozí rovnice pro $x_i < -\tau$.

- V bodech, kde $g(y)$ není diferencovatelná, platí $y_i = 0$ a protože pro $|x_i| > \tau$ již máme příslušnou složku řešení y_i definovanou, platí $y_i = 0$ pro $x_i \in \langle -\tau, \tau \rangle$.

Všechny předchozí možnosti lze sloučit právě do podoby tvrzení. \square

Poznámka. Funkce $\text{prox}_{\tau\|\cdot\|_1}(x)$ se nazývá *měkké prahování s prahem τ* (v anglické literatuře *soft thresholding*) a budeme ji dále zkráceně značit $\text{soft}_\tau(x)$.

Definice 18. *Nechť C je neprázdná podmnožina \mathbb{R}^N . Funkci*

$$\iota_C(x) = \begin{cases} 0, & \text{pro } x \in C \\ \infty, & \text{pro } x \notin C \end{cases} \quad (4.19)$$

nazýváme indikátorovou funkcí množiny C .

Nyní chceme vektor x ortogonálně promítnout na množinu $C = \{z : D^r z = y^r\}$, ekvivalentně hledáme řešení úlohy

$$\text{prox}_{\iota_C}(x). \quad (4.20)$$

Lemma 1. [5] *Označme $A = \{z : Dz = 0\}$. Pak*

$$\text{prox}_{\iota_A}(x) = x - D^+ D x = (I - D^+ D)x. \quad (4.21)$$

Důkaz. Z definice proximálního operátoru vidíme, že hledáme takový vektor z , který splňuje $Dz = 0$ a současně $\|x - z\|_2^2 = \|r\|_2^2$ je minimální.

Vyjádřením $x = z + r$ a násobením maticí D dostaneme

$$Dx = Dz + Dr = Dr. \quad (4.22)$$

Tuto rovnici můžeme chápat jako soustavu s maticí D , neznámou r a pravou stranou Dx . Z vlastností pseudoinverze víme, že tato soustava má řešení $D^+(Dx)$ a navíc se jedná o řešení s minimálním kvadrátem normy. Dostáváme tedy požadované minimální $r = D^+(Dx)$ a vyjádřením $z = x - r$ je lemma dokázáno. \square

Lemma 2. *Označme $B = \{z : Dz = y\}$. Pak*

$$\text{prox}_{\iota_B}(x) = x - D^+(Dx - y) = (I - D^+ D)x + D^+ y. \quad (4.23)$$

Důkaz. V důkazu postupujeme stejně jako v důkazu předchozího lemmatu. Rovnice (4.22) bude ve tvaru

$$Dx = Dz + Dr = y + Dr. \quad (4.24)$$

Řešíme tedy soustavu s pravou stranou $Dx - y$. Hledaným minimálním řešením je $r = D^+(Dx - y)$ a vyjádřením $y = x - r$ je lemma dokázáno. \square

Následující lemma i jeho důkaz jsou převzaty z článku [15]. Významem těchto úprav je to, že v pozdějších výpočtech proximálního operátoru se za uvedených podmínek vyhneme výpočetně náročnému hledání pseudoinverze matice.

Lemma 3. *Nechť D je Parsevalův frame. Potom pro pseudoinverzi redukovaného slovníku D^r platí*

$$(D^r)^+ = D^*(I^r)^T. \quad (4.25)$$

Důkaz. Díky vlastnosti pseudoinverze $A^+ = A^*(AA^*)^+$ máme

$$(D^r)^+ = (D^r)^*(D^r(D^r)^*)^+. \quad (4.26)$$

Dále využijeme vlastnosti hermitovské transpozice součinu – protože $D^r = I^r D$, platí $(D^r)^* = D^*(I^r)^* = D^*(I^r)^T$. Dostáváme tedy rovnici

$$(D^r)^+ = (D^r)^* \left(I^r D D^* (I^r)^T \right)^+. \quad (4.27)$$

Protože D je Parsevalův frame, platí $DD^* = I$ a tudíž $\left(I^r D D^* (I^r)^T \right)^+$ je opět jednotková matice. Dostáváme tedy

$$(D^r)^+ = (D^r)^* I = (I^r D)^* = D^*(I^r)^T. \quad (4.28)$$

□

Hledaný proximální operátor vyjádříme v následujícím tvrzení:

Tvrzení 5. *Označme $C = \{z : D^r z = y^r\}$. Pak*

$$\text{prox}_{\iota_C}(x) = x - D^*(I^r)^T I^r D x + D^*(I^r)^T y^r. \quad (4.29)$$

Důkaz. Tvrzení plyne přímo z lemmatu 2, kde místo D uvažujeme D^r , místo y máme y^r a pro výpočet příslušné pseudoinverze využijeme lemma 3. □

4.2.4. Návrh optimalizačního algoritmu

Pro funkce f a g zavedené v (4.6) a (4.9) jsou splněny podmínky konvergence DR algoritmu, tzn. $\text{dom}(f) \cap \text{dom}(g) \neq \emptyset$ a $\lim_{\|x\| \rightarrow \infty} f(x) + g(x) = \infty$. Podle předchozí části jsou příslušné proximální operátory následující (předpokládáme použití Parsevalova framu):

$$\text{prox}_{\tau f}(x) = x - D^*(I^r)^T I^r D x + D^*(I^r)^T y^r, \quad (4.30)$$

$$\text{prox}_{\tau g}(x) = \text{soft}_{\tau}(x). \quad (4.31)$$

Operátor $\text{prox}_f = \text{prox}_{\tau \iota_C}$ se nijak neliší od výše uvedeného prox_{ι_C} , protože indikátorová funkce se násobením kladnou konstantou nijak nezmění. Násobení maticí D značí syntézu užitím zvoleného framu, násobení maticí D^* značí analýzu.

DR algoritmus pro řešený problém

Následující algoritmus přímo vychází z obecného DR algoritmu 2, pouze za proximální operátory byla dosazena jejich konkrétní vyjádření a vyjádření y_{n+1} bylo zjednodušeno.

Algoritmus 3: DR algoritmus pro zaplnění chybějícího úseku signálu

```

1 zvolíme  $\varepsilon \in (0, 1), \tau > 0, y_0 \in \mathbb{R}^N$ 
2 for  $n = 0, 1, 2, \dots$  do
3    $x_n = \text{soft}_\tau(y_n)$ 
4   zvolíme  $\lambda_n \in \langle \varepsilon, 2 - \varepsilon \rangle$ 
5    $y_{n+1} = (1 - \lambda_n) y_n + \lambda_n \left( x_n - D^* (I^r)^T I^r D (2x_n - y_n) + D^* (I^r)^T y^r \right)$ 
6 end
7  $x = x_n - D^* (I^r)^T I^r D x_n + D^* (I^r)^T y^r$ 

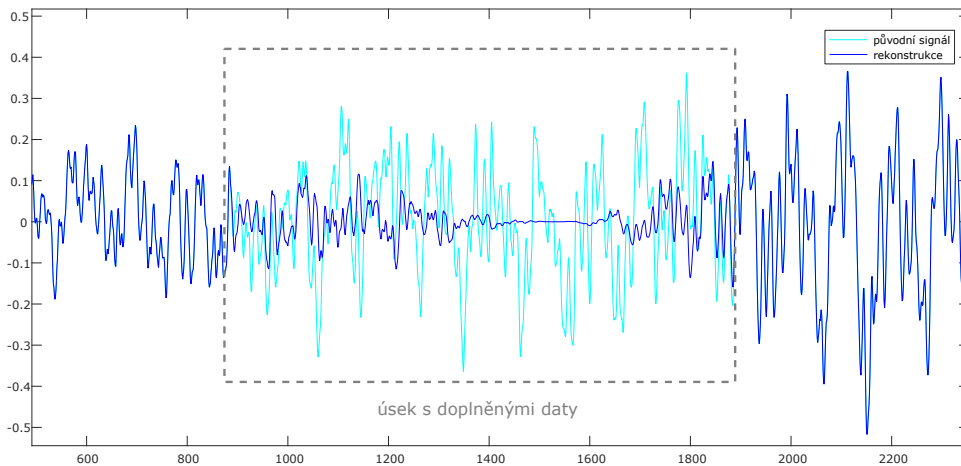
```

Iterace končí ve chvíli, kdy hodnota kritéria $\frac{\|x_{n+1} - x_n\|}{\|x_n\|}$ je menší než předem zvolená mez přesnosti, nebo je dosažen předem daný maximální počet iterací. V projekci v posledním kroku každé iterace také vidíme, že výpočty $(I^r)^T I^r$ a $D^* (I^r)^T y^r$ lze provést pouze jednou a není třeba je v každé iteraci opakovat.

Jako poslední krok algoritmu byla přidána projekce $x = \text{prox}_{\tau f} x_n$, abychom při konečném počtu iterací měli zaručeno splnění podmínky $D^r x = y^r$.

4.2.5. Maximální řešitelná délka díry

Při experimentech bylo zjištěno, že od určité délky díry se v rekonstruovaném signálu objeví úsek obsahující pouze nulové vzorky. Tato část se věnuje teoretické maximální délce díry, kterou může zaplnit algoritmus, jenž využívá Gaborovy či podobné transformace pracující s okny, aniž by se po rekonstrukci takový nulový úsek objevil. Vycházíme ze zadaných parametrů oken framu a počátečního indexu díry v signálu a počítáme maximální délku díry h_{\max} , kterou může výše specifikovaný algoritmus zpracovat.



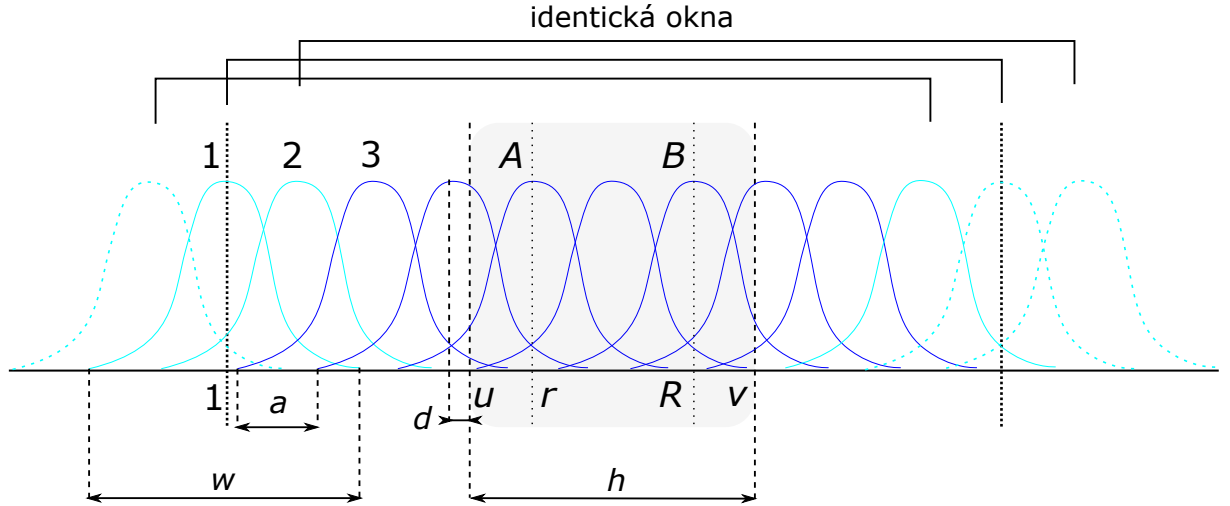
Obrázek 4.1: Příklad výskytu nulového úseku po rekonstrukci příliš dlouhé díry v signálu.

Obrázek 4.2 a značení vychází z článku [14], který se zabýval související restrikcí signálu – cílem bylo v závislosti na poloze díry zkrátit signál na takové okolí díry, aby

rekonstrukce proběhla bez negativního vlivu periodizace krajních oken framu (v obrázku se jedná o modře znázorněná okna). Protože index posledního prvku a číslo $h = h_{\max}$ je cílem vypočítat, jsou v obrázku pouze pro úplnost.

Odvozený model pro maximální délku díry je pouze teoretický – jak bude ukázáno v následující kapitole, v této práci zvolený algoritmus pracuje pouze pro díry znatelně kratší.

Popis situace



Obrázek 4.2: Vizualizace posloupnosti oken Gaborova framu v okolí díry.

- a posun oken framu
- w délka okna framu, pro zjednodušení budeme předpokládat, že w je sudé
- u index prvního prvku díry
- v index posledního prvku díry
- h délka díry v signálu
- A index posledního okna zasahujícího ještě do levé části signálu
- B index prvního okna zasahujícího již do pravé části signálu
- r centrální index okna A , $r = 1 + (A - 1)a$
- R centrální index okna B , $R = 1 + (B - 1)a$
- d vzdálenost začátku díry od nejbližšího předchozího centrálního prvku okna

Délkou díry v signálu rozumíme počet prvků, o kterých nemáme žádnou informaci. Lze tedy psát $h = v - u + 1$. Centrálním indexem okna 1 rozumíme prvek na pozici 1, konec tohoto okna je tedy na pozici $\frac{w}{2}$ (protože předpokládáme w sudé, je tento zlomek celočíselný). Obecně okno $n + 1$ má svůj centrální index $1 + n \cdot a$, první prvek tohoto okna je na pozici $1 + n \cdot a - \frac{w}{2}$ a poslední na pozici $n \cdot a + \frac{w}{2}$. Protože $u = (1 + n \cdot a) + d$ pro nějaké $n \in \mathbb{N}$ a $d < a$, platí $u - 1 = n \cdot a + d$ a tudíž $d = (u - 1) \bmod a$.

Nulové vzorky po rekonstrukci poškozeného signálu vzniknou tehdy, pokud nemá žádné překrytí poslední okno zasahující ještě do levé části signálu (A) a první okno zasahující již do pravé části signálu (B). Abychom zaručili, že se okna A a B budou překrývat alespoň jedním prvkem, musí platit

$$r - 1 + \frac{w}{2} \geq R - \frac{w}{2}. \quad (4.32)$$

Tuto nerovnici upravíme do přehlednějšího tvaru

$$r + w > R. \quad (4.33)$$

Díky tomu, že počítáme pouze s celými čísly (jedná se o indexy prvků vektoru), mohli jsme nerovnici (4.32) snadno upravit do tvaru s ostrou nerovností.

Dále budeme požadovat, aby okna A a B zasahovala do neporušené části signálu každé alespoň jedním vzorkem – v opačném případě by nenesla žádnou informaci o hledané rekonstrukci.

Nalezení okna A

Hledáme největší index r , který splňuje podmínky

$$u > r - \frac{w}{2} \quad \wedge \quad r = 1 + k \cdot a \quad \text{pro nějaké } k \in \mathbb{N}. \quad (4.34)$$

Hledaným parametrem je nyní číslo k . Dosazením druhé podmínky do první dostaneme nerovnici

$$1 + k \cdot a < u + \frac{w}{2}. \quad (4.35)$$

Pro příští úpravy bude výhodnější tvar s neostrou nerovností

$$k \cdot a \leq u + \frac{w}{2} - 2, \quad (4.36)$$

na něžž můžeme rovnici (4.35) převést díky tomu, že všechny členy v ní vystupující jsou přirozená čísla. Podmínku pro k tedy dostaneme ve tvaru

$$k \leq \frac{u + \frac{w}{2} - 2}{a}. \quad (4.37)$$

Protože hledáme největší $k \in \mathbb{N}$ splňující předchozí nerovnici, dostaneme řešení

$$k = \left\lfloor \frac{u + \frac{w}{2} - 2}{a} \right\rfloor. \quad (4.38)$$

Centrální index okna A tedy můžeme vyjádřit jako

$$r = 1 + \left\lfloor \frac{u + \frac{w}{2} - 2}{a} \right\rfloor \cdot a. \quad (4.39)$$

Nalezení okna B

Okno B bude poslední okno, které se bude ještě překrývat alespoň jedním prvkem s oknem A . Pro centrální indexy oken A a B musí platit

$$R = r + m \cdot a \quad \text{pro nějaké } m \in \mathbb{N}. \quad (4.40)$$

Navíc pro jednoduchost předpokládejme, že délka okna je celočíselným násobkem posunu mezi okny, což formulujeme ve tvaru $a = \frac{w}{n}$ pro dané $n \in \mathbb{N}$. Dosazením do (4.33) dostaneme

$$r + n \cdot a > R = r + m \cdot a. \quad (4.41)$$

Z předchozí nerovnice hned vidíme, že hledané přirozené číslo m , které vyjadřuje posun mezi okny A a B , musí splňovat podmínku $m < n$. Protože chceme okny A a B pokrýt co nejdelší díru, hledáme m co největší, tudíž volíme $m = n - 1$. Dohromady tedy dostáváme

$$R = r + (n - 1) a = r + w - a = 1 + \left\lfloor \frac{u + \frac{w}{2} - 2}{a} \right\rfloor \cdot a + w - a. \quad (4.42)$$

Určení maximální délky díry

Pro dané parametry framu a začátek díry už máme nalezena příslušná okna A a B . Nyní hledáme index posledního prvku díry tak, aby bylo (za předpokladů uvedených v úvodu této části) možné její zaplnění.

Poslední index díry musí být menší než index posledního prvku okna B , který je na pozici $R + \frac{w}{2} - 1$. Index posledního prvku díry tedy může být nejvýše $R + \frac{w}{2} - 2$. Maximální délka díry potom bude

$$h_{\max} = \left(R + \frac{w}{2} - 2\right) - u + 1. \quad (4.43)$$

Číslem h_{\max} rozumíme počet prvků, které mohou do díry spadat. Pokud do předchozího vyjádření dosadíme vztah (4.42) pro R , dostaneme

$$h_{\max} = \left(1 + \left\lfloor \frac{u + \frac{w}{2} - 2}{a} \right\rfloor \cdot a + w - a + \frac{w}{2} - 2\right) - u + 1, \quad (4.44)$$

což lze upravit na

$$h_{\max} = \frac{3w}{2} + a \cdot \left(\left\lfloor \frac{u + \frac{w}{2} - 2}{a} \right\rfloor - 1\right) - u. \quad (4.45)$$

Na závěr ukážeme, že h_{\max} není závislé na poloze okna A vzhledem k začátku signálu, ale pouze na parametru d , který je znázorněný na obrázku 4.2. Vyjádříme $u = (1 + n \cdot a) + d$ pro nějaké $n \in \mathbb{N}$ a dosadíme do předchozí rovnice:

$$h_{\max} = \frac{3w}{2} + a \cdot \left(\left\lfloor \frac{n \cdot a + d + \frac{w}{2} - 1}{a} \right\rfloor - 1\right) - 1 - n \cdot a - d = \quad (4.46)$$

$$= \frac{3w}{2} + a \cdot \left(\left\lfloor n + \frac{d + \frac{w}{2} - 1}{a} \right\rfloor - 1\right) - 1 - n \cdot a - d. \quad (4.47)$$

Protože n je přirozené číslo, můžeme ho vyjmout mimo dolní celou část a po roznásobení závorky číslem a se tento člen odečte s členem $-n \cdot a$. Dostaneme tedy konečný tvar

$$h_{\max} = \frac{3w}{2} + a \cdot \left(\left\lfloor \frac{d + \frac{w}{2} - 1}{a} \right\rfloor - 1\right) - 1 - d. \quad (4.48)$$

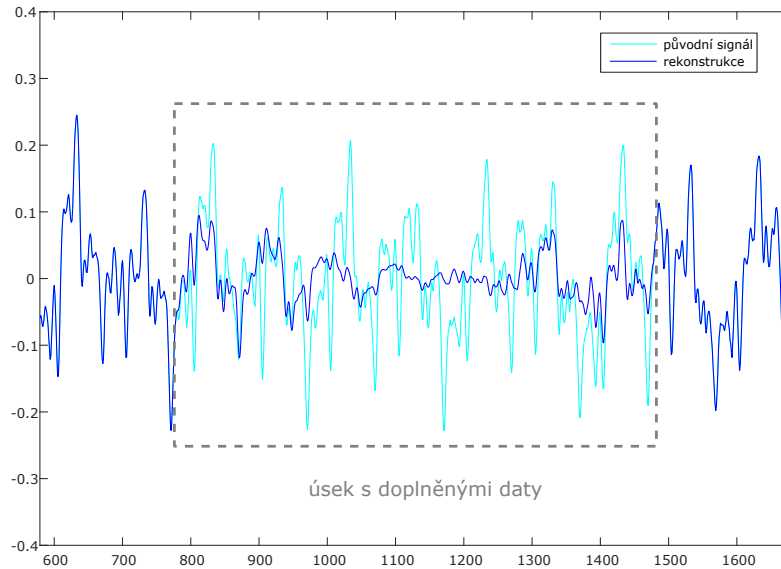
Z této rovnice vidíme, že delší díry v signálu budou lépe pokryty framy s větší délkou okna w a s malým posunem a . Pokud nebudeme uvažovat dolní celou část, dostaneme úpravou poslední rovnice maximální délku díry $h_{\max} = 2w - a - 2$. Protože použitím dolní celé části se může h_{\max} jedině snížit, vyjadřuje tento výsledek horní závorku možných výsledků pro hodnoty parametru d , tedy

$$h_{\max} \leq 2w - a - 2. \quad (4.49)$$

Z extrémního případu $a = 1$ tedy vidíme, že signál s dírou délkou větší než $h = 2w - 3$ nemůže být v našem modelu žádným Gaborovým framem rekonstruován, aniž by se po rekonstrukci vyskytl úsek pouze nulových vzorků.

4.2.6. Kompenzace poklesu energie rekonstruovaného signálu

Z obrázku 4.3 je zřejmé, že amplituda rekonstruovaného signálu (znázorněn modře) je v doplněném úseku výrazně nižší oproti amplitudě neporušeného signálu (znázorněn tyrkysově), což odpovídá nižší energii, jak byla zavedena v definici 6. Navíc amplituda rekonstrukce klesá směrem dovnitř doplněného úseku.



Obrázek 4.3: Ukázka poklesu energie rekonstruovaného signálu v místě doplněných vzorků.

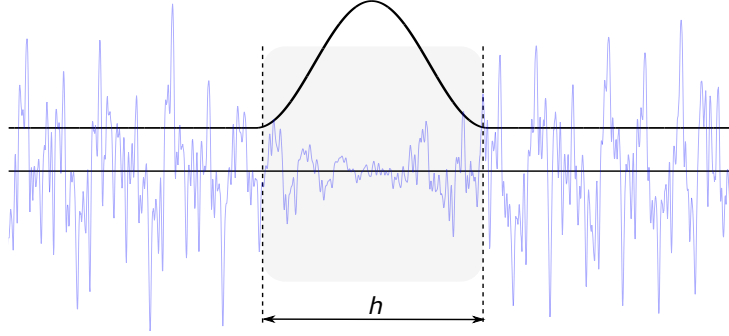
Poznámka. Popsané porovnání signálů je možné díky tomu, že jsme použili primárně nepoškozený audiosignál, ve kterém jsme úsek chybějících vzorků uměle vytvořili.

Tento pokles amplitudy je pravděpodobně důsledkem řešení relaxované úlohy (4.8) místo původní (4.3). Minimalizací ℓ_1 -normy při hledání řídké reprezentace totiž upřednostňujeme řešení, která jsou nejen řídká, ale navíc jejich nenulové koeficienty jsou blíží nule. V neporušené části signálu je tento fakt nedůležitý, neboť podmínkou $D^r x = y^r$ v úloze (4.8) je zajištěna rovnost původního a rekonstruovaného signálu. V místě doplněných vzorků se však upřednostnění koeficientů s menší absolutní hodnotou projeví jako pokles amplitudy signálu a tudíž i jeho energie.

V této části tedy navrhneme metody, jak pokles energie po zaplnění chybějících vzorků kompenzovat. Předpokládáme přitom, že nebyla dosažena maximální délka díry odvozená v předchozí části, tudíž máme k dispozici signál bez úseku nulových vzorků.

Kompenzace poklesu energie ve výsledném signálu

Tato heuristická metoda je jednoduše implementovatelná, relevantnost jejího výsledku je však spekulativní. Základním principem je vynásobení výsledného signálu dle postupu v části 4.2.1 po složkách vhodnou funkcí, která signál mimo díru zachová a doplněným vzorkům zvětší amplitudu, čímž kompenzuje pokles energie. Ukázka takové funkce je na obrázku 4.4.



Obrázek 4.4: Návrh kompenzační funkce pro rekonstruovaný signál se zaplněnými vzorky v označeném úseku délky h . Zvolenou funkcí je jedna perioda funkce kosinus, která na hranici úseku doplněných dat hladce navazuje na konstantní funkci s hodnotou 1.

Problémem metody je určení kompenzační funkce tak, aby odrážela příčinu poklesu energie. Proto je tato metoda implementována se zvolenou hladkou kompenzační funkcí pouze jako závěrečný krok některé z pokročilejších metod představených níže.

Kompenzace poklesu energie v průběhu minimalizace

Nyní se pokusíme navrhnout postup, který by bránil poklesu energie již v průběhu algoritmu pro doplnění dat. Jak bylo uvedeno v úvodu této části, dosavadní algoritmus 3 hledá řešení s minimální ℓ_1 -normou, což realizuje v každé iteraci měkké prahování – je-li absolutní hodnota složky reprezentace větší než zvolený práh, je tato absolutní hodnota o hodnotu prahu zmenšena. Protože chceme regulovat snižování koeficientů příslušejících atomům, které přispívají k zaplnění chybějících vzorků (a tím bránit poklesu energie ve výsledné rekonstrukci), nahradíme měkké prahování s konstantním prahem τ prahováním s prahem $\tau \in \mathbb{R}^M$, kde M značí počet složek hledané reprezentace. V následujícím tvrzení tedy upravíme měkké prahování pro případ, kdy jeho parametrem je vektor.

Tvrzení 6. Pro libovolný vektor $\tau \in \mathbb{R}^M$, $\tau_i > 0$, $i = 0, 1, \dots, M-1$, zavedme čtvercovou matici T řádu M , která má na diagonále prvky vektoru τ a prvky mimo diagonálu má nulové (značíme $T = \text{diag } \tau$). Pak

$$\text{prox}_{\|T \cdot\|_1}(x) = (\text{sgn}(x_i) \cdot \max\{|x_i| - \tau_i, 0\})_{i=0}^{M-1}. \quad (4.50)$$

Důkaz. Důkaz je obdobný jako u tvrzení 4, pouze místo konstanty τ uvažujeme v jednotlivých složkách příslušnou složku τ_i . \square

Poznámka. Značením $\|T \cdot\|_1$ rozumíme modifikovanou ℓ_1 -normu danou předpisem

$$\|Tx\|_1 = \sum_{i=0}^{M-1} |T_{i,i}x_i| = \sum_{i=0}^{M-1} |\tau_i x_i|, \quad (4.51)$$

což je právě tvar, pro který lze analogicky k důkazu tvrzení 4 odvodit tvrzení 6.

Můžeme tedy říci, že pro zvolenou matici $T = \text{diag}(\tau)$ řešíme místo původní (relaxované) úlohy 4.8 modifikovanou optimalizační úlohu

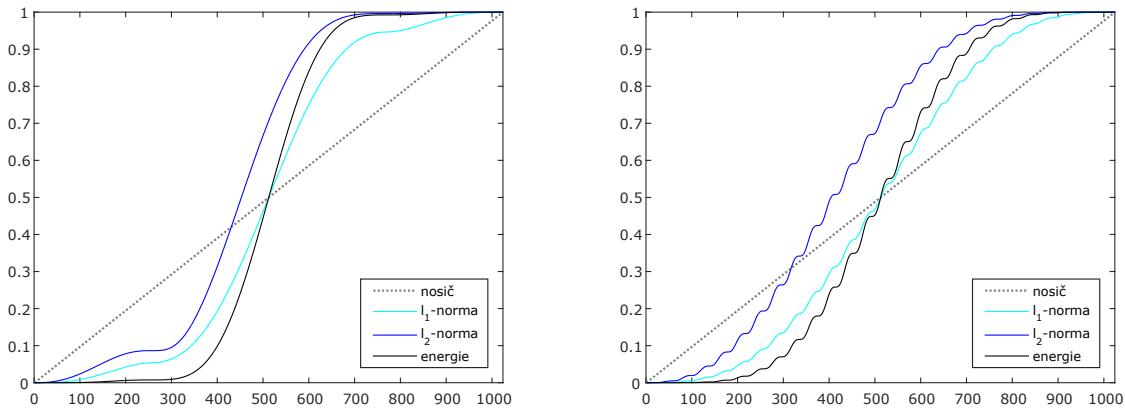
$$\arg \min_x \|Tx\|_1 \text{ vzhledem k } D^T x = y^T. \quad (4.52)$$

Pro určení vektoru τ nyní potřebujeme každému atomu zvoleného slovníku přiřadit číslo $w_i \in \langle 0, 1 \rangle$, které budeme nazývat *váha atomu* a vektor $(w_i)_{i=0}^{M-1}$ označíme *váhovací funkce*. Váhy w_i poté určují, kolik informace o neporušené části signálu jednotlivé atomy nesou. Budeme je tedy určovat tak, aby atom, jehož nosič nezasahuje do úseku chybějících vzorků, měl váhu 1. Naopak u atomu, jehož nosič je celý součástí úseku chybějících vzorků, požadujeme váhu 0. Pro funkčnost metody předpokládáme, že složky reprezentace, kterým přiřadíme váhu 0, budou v rámci DR algoritmu vynulovány (což zajistíme vhodnou volbou výchozího bodu y_0 a projekcí v kroku 5 algoritmu 3), čímž se vyhneme nepřipustnému prahování s nulovým prahem.

Označme nyní d_i , $i = 0, 1, \dots, M-1$, jednotlivé atomy zvoleného slovníku a d_i^r příslušné části atomů po odebrání složek odpovídajících chybějícím vzorkům původního signálu (odebráním zde myslíme zkrácení vektoru). Návrhy pro určení váhovací funkce jsou pak následující:

$$\begin{array}{lll} \text{I} & w_i = 1, & \text{II} \quad w_i = \frac{|\text{supp}(d_i^r)|}{|\text{supp}(d_i)|}, \quad \text{III} \quad w_i = \frac{\|d_i^r\|_1}{\|d_i\|_1}, \\ \text{IV} & w_i = \frac{\|d_i^r\|_2}{\|d_i\|_2}, & \text{V} \quad w_i = \frac{\|d_i^r\|_2^2}{\|d_i\|_2^2}. \end{array}$$

První případ je zde pro úplnost, váhy jsou zde konstantní, jedná se tedy o původní podobu algoritmu bez váhování. V druhém případě předpokládáme, že význam jednotlivých atomů pro rekonstrukci chybějících vzorků závisí pouze na tom, jaká část nosiče zasahuje do úseku díry. Třetí případ pak uvažuje ℓ_1 -normu, čtvrtý ℓ_2 -normu a poslední energii části atomu odpovídající neporušenému signálu vzhledem k atomu celému.



Obrázek 4.5: Kumulační křivky pro navržené váhy. Zvoleným slovníkem je Gaborův frame generovaný Hannovým oknem φ délky 1024 vzorků. Vlevo jsou průběhy pro atom $\pi_{0,2}(\varphi)$, vpravo pro atom $\pi_{0,500}(\varphi)$.

Obrázek 4.5 ilustruje, že předchozím rozdělením skutečně získáme výrazně odlišné váhovací funkce. Pro dva různé atomy framu generovaného Hannovým oknem jsou vykresleny kumulační křivky pro jednotlivé návrhy váhovací funkce. Kumulační křivkou zde rozumíme pro daný atom d délky N zobrazení

$$n \mapsto w(n), \quad n = 0, 1, \dots, N-1, \quad (4.53)$$

kde $w(n)$ je zvolená váha, při jejímž výpočtu uvažujeme (dle předchozího značení) vektory

$$d_i = d, \quad d_i^r = (d_0, d_1, d_2, \dots, d_n). \quad (4.54)$$

Navrhli jsme tedy čtyři způsoby generování váhovací funkce. Předem však nelze odhadnout, kterým způsobem dosáhneme při rekonstrukci nejlepších výsledků. Výhody a nevýhody jednotlivých způsobů pro použití při rekonstrukci konkrétního signálu proto budou blíže rozebrány v části 5.

4.2.7. Návrh implementace algoritmu

Algoritmus pro doplnění chybějícího úseku audiosignálu je implementován v softwaru MATLAB s využitím toolboxu pro krátkodobou Fourierovu analýzu LTFAT [12] (z anglického *The Large Time-Frequency Analysis Toolbox*, dostupné online: <http://ltfat.sourceforge.net/>). Celý postup můžeme shrnout v algoritmu 4.

Algoritmus 4: Algoritmus pro zaplnění chybějícího úseku signálu

- 1 Volba (těsného) Gaborova framu D – typ okna a délka jeho nosiče w , posun a a počet kmitočtových pásem M ,
 - 2 určení začátku a konce úseku chybějících vzorků a jeho délky,
 - 3 výpočet váhovací funkce dle některého z postupů uvedeného v části 4.2.6,
 - 4 vygenerování matice I^r potřebné pro práci s redukováným slovníkem D^r ,
 - 5 určení výchozího bodu pro hledání řídké reprezentace analýzou signálu y^r , ve kterém byly na místa chybějících vzorků vloženy vzorky nulové,
 - 6 nalezení řídké reprezentace pomocí algoritmu 3, ve kterém volíme parametr τ s ohledem na krok 3, případně jako kladnou konstantu,
 - 7 syntéza rekonstruovaného signálu z koeficientů nalezených v kroku 6,
 - 8 možnost závěrečné kompenzace možného poklesu energie v rekonstruovaném signálu postupem navrženým v části 4.2.6.
-

Pro urychlení řadíme za druhý krok zkrácení signálu dle algoritmu prezentovaného v článku [14]. Toolbox LTFAT navíc obsahuje funkce pro analýzu i syntézu signálu pomocí zvoleného framu, vyhneme se tedy častému maticovému násobení.

5. Experimenty a vyhodnocení

V této kapitole nejdříve revidujeme odvozené vztahy pro maximální délku díry z části 4.2.5. Tento úsek je zařazen pouze jako ukázka; hlubší statistický výzkum, který by umožnil stanovení níže zavedených parametrů pro libovolný signál, je nad rámec této práce.

V další části se pak věnujeme rozsáhlejšímu experimentu, na kterém poté zhodnotíme kvalitu navrženého algoritmu.

5.1. Maximální řešitelná délka díry

Jak již bylo výše uvedeno, vzorec (4.48) pro maximální délku díry platí pouze teoreticky, pokud stačí, aby přesah oken do neporušené části signálu byl jeden vzorek. Pokud ovšem řešíme problém rekonstrukce díry v signálu DR algoritmem, tento předpoklad není splněn. Při rekonstrukci se totiž koeficienty příslušející oknům s malým přesahem mimo díru vynulují. Krátké úseky neporušené části signálu, které přímo sousedí s dírou, jsou pak s dostatečnou přesností rekonstruovány i bez využití těchto oken – algoritmus upřednostní atomy oken, jejichž vliv na signál je v příslušných místech výrazně vyšší.

Situaci ilustruje následující experiment. Do výpočtu pro maximální délku díry zařadíme parametry par_A a par_B , které označují, jaká část nosiče oken A a B musí zasahovat do neporušeného signálu, aby se tato okna v rekonstrukci projevila.

Nerovnice (4.36) potom bude mít tvar

$$k \cdot a \leq u + \frac{w}{2} - (1 + par_A) \quad (5.1)$$

a hledané číslo r bude

$$r = 1 + \left\lfloor \frac{u + \frac{w}{2} - (1 + par_A)}{a} \right\rfloor \cdot a. \quad (5.2)$$

Pro index R bude platit (na základě (4.42))

$$R = 1 + \left\lfloor \frac{u + \frac{w}{2} - (1 + par_A)}{a} \right\rfloor \cdot a + w - a. \quad (5.3)$$

Parametr par_B se projeví ve výpočtu posledního indexu díry, protože požadujeme, aby byl alespoň o par_B nižší než index posledního prvku okna B . Protože tento prvek je na pozici $R + \frac{w}{2} - 1$, může být poslední index díry nejvýše $R + \frac{w}{2} - (1 + par_B)$.

Po těchto úpravách dostaneme vzorec pro maximální délku díry ve tvaru

$$h_{\max} = \left(1 + \left\lfloor \frac{u + \frac{w}{2} - (1 + par_A)}{a} \right\rfloor \cdot a + w - a + \frac{w}{2} - (1 + par_B) \right) - u + 1, \quad (5.4)$$

který obdobně jako v části 4.2.5 upravíme na tvar

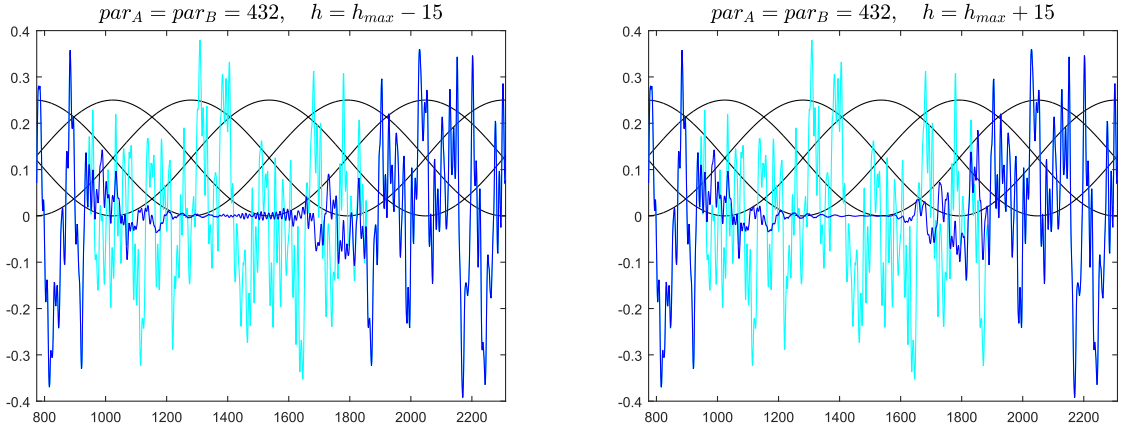
$$h_{\max} = \frac{3w}{2} + a \cdot \left(\left\lfloor \frac{u + \frac{w}{2} - (1 + par_A)}{a} \right\rfloor - 1 \right) - u + 1 - par_B \quad (5.5)$$

a závěrečným zavedením parametru $d = (u - 1) \bmod a$ dostaneme výsledek

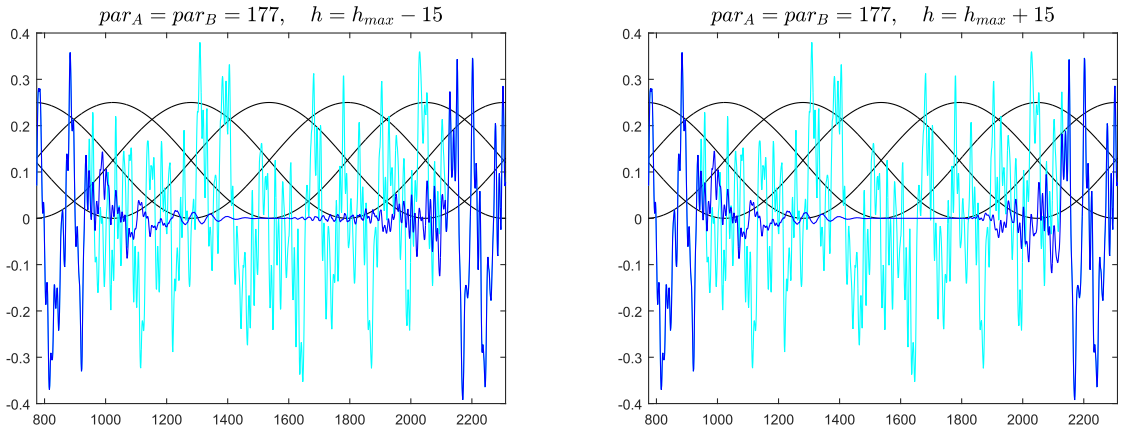
$$h_{\max} = \frac{3w}{2} + a \cdot \left(\left\lfloor \frac{d + \frac{w}{2} - par_A}{a} \right\rfloor - 1 \right) - d - par_B. \quad (5.6)$$

Alternativně lze tedy říci, že teoretická délka díry v rovnici (4.48) je specifickým případem rovnice (5.6), jestliže položíme $par_A = par_B = 1$.

V realizovaném experimentu byl použit signál lidové hudby s vzorkovací frekvencí 44,1 kHz a s uměle vytvořenou dírou délky $h_{\max} \pm 15$ vzorků, kde h_{\max} bylo určeno podle vzorce (5.6) pro zvolenou hodnotu $par_A = par_B = C$. Použitý frame byl těsný frame generovaný Hannovým oknem délky $w = 1024$ (odpovídá cca 23 ms), posunem mezi okny $a = 256$ a $M = 1024$ rovnoměrně rozloženými kmitočtovými pásmy. Následující obrázky zobrazují v časové oblasti původní a rekonstruovaný signál v závislosti na parametrech par_A a par_B . Pro $par_A = par_B = 432$ vzorků je $h_{\max} = 928$, pro $par_A = par_B = 177$ vzorků je $h_{\max} = 1183$. V ideálním případě ($par_A = par_B = 1$) je $h_{\max} = 1615$.



Obrázek 5.1: Nastavení parametrů, pro které se objeví pokles energie rekonstruovaného signálu na nulu, $h_{\max} = 928$ vzorků. Tyrkysově je vykreslen původní signál, modře signál rekonstruovaný. Černě jsou symbolicky znázorněna jednotlivá okna framu.



Obrázek 5.2: Nastavení parametrů, pro které se objeví dlouhý nulový úsek v rekonstruovaném signálu, $h_{\max} = 1183$ vzorků. Tyrkysově je vykreslen původní signál, modře signál rekonstruovaný. Černě jsou symbolicky znázorněna jednotlivá okna framu.

5.2. Algoritmus pro doplňování chybějících úseků audiosignálu a možnosti kompenzace poklesu energie rekonstruovaného signálu

5.2.1. Návrh experimentu

Pro testování byly zvoleny audiosignály

- (a) Glockenspiel (zvonkohra),
- (b) housle,
- (c) housle a viola,
- (d) hudecká muzika,
- (e) cimbálová muzika,
- (f) rocková skupina.

Z každého signálu byl vybrán vzorek délky 6 s, který byl na hranicích vyhlazen, čímž vznikl plynulý náběh a útlum amplitudy signálu (fade-in a fade-out) délky 0,3 s. Vytvořené vzorky se výrazně liší řídkostí, což ilustrují spektrogramy na obrázku 5.3 – vzorky jsou zde řazeny od nejřidšího po nejméně řídský signál. Protože náš algoritmus je založený právě na řídkosti, očekáváme, že pro řidší signály dosáhneme lepších výsledků.

V každém ze vzorků poté vytvoříme jednu díru zvolených parametrů. Délku díry volíme s ohledem na výsledek v části o maximální řešitelné délce díry v signálu, čímž ji omezíme shora. Vzhledem ke skutečnosti nastíněné v části 5.1 však nebudeme provádět větší počet testů pro délky díry blížíící se této maximální hodnotě. Současně nemá význam testovat algoritmus pro díry příliš krátké, budeme tedy volit délky v rozmezí od počtu vzorků odpovídajících 1 ms po hodnotu v nerovnici (4.49). Vhodně rozložené délky generujeme pomocí funkce

$$h = \alpha \cdot 2^{x/10} + x + \beta, \quad (5.7)$$

kde h označuje počet chybějících vzorků, x je index realizace experimentu (od 1 do 100) a parametry α a β určíme tak, aby platilo

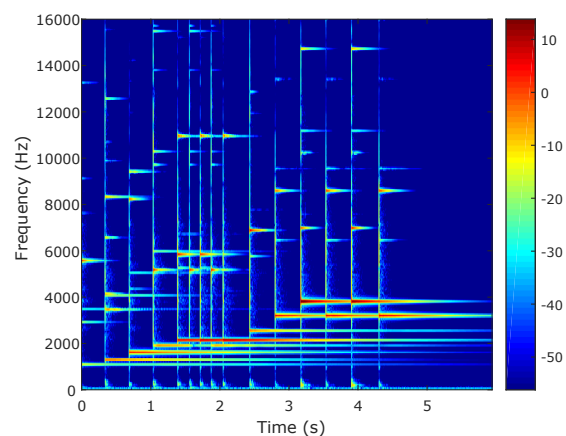
$$h(1) = 0,001 \cdot f_s, \quad (5.8)$$

$$h(100) = 2w - a - 2. \quad (5.9)$$

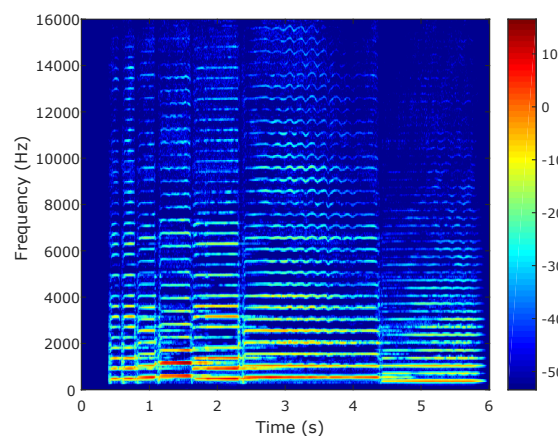
První podmínka zajišťuje, že nejkratší délka díry odpovídá 1 ms (f_s značí vzorkovací frekvenci). Druhá podmínka je pak omezením maximální délky díry, která je teoreticky zaplnitelná bez výskytu úseku pouze nulových vzorků s využitím Gaborova framu s délkou okna w a posunem a . Pro lepší možnost porovnání nebudeme délky díry měnit při použití jiného framu, volíme tedy pevnou hodnotu $w = 1024$ vzorků a $a = 256$. Dosazením těchto podmínek dostaneme

$$\alpha = \frac{h(100) - h(1) - 99}{2^{10} - 2^{1/10}}, \quad \beta = h(1) - \alpha \cdot 2^{1/10} - 1. \quad (5.10)$$

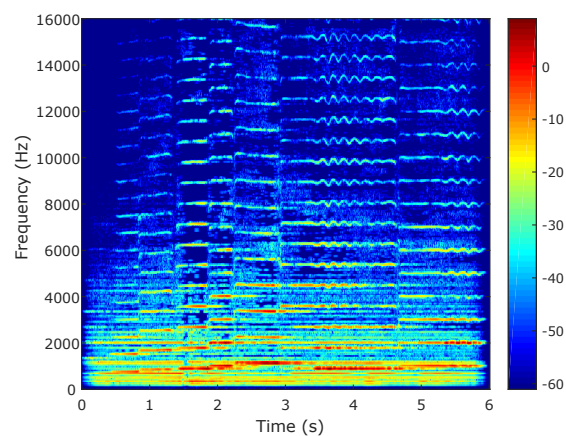
Začátek díry zvolíme náhodně v úseku od 0,5 s do 5,5 s (s využitím funkce `randi` v MATLABu). Tímto se vyhneme vlivu upraveného začátku a konce signálu. V případě, že bychom uvažovali celou délku vzorku, mohly by být výsledky ovlivněny výskytem díry,



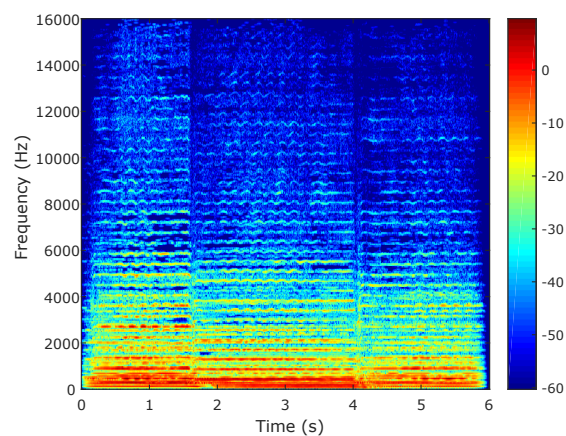
(a) Glockenspiel (zvonkohra)



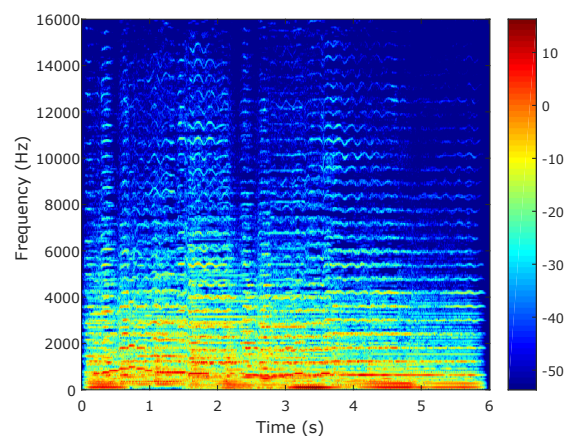
(b) housle



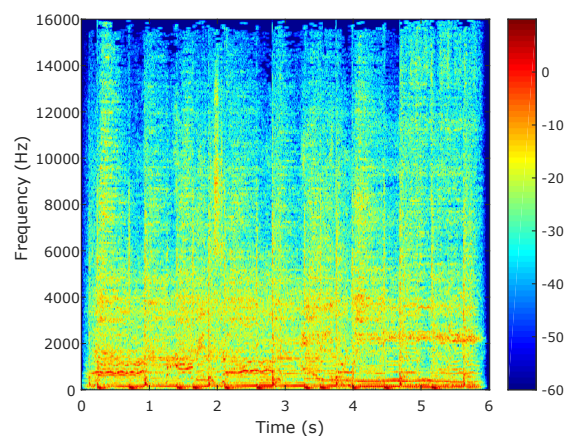
(c) housle a viola



(d) hudecká muzika (dvoje housle, viola, kontrabas)



(e) cimbálová muzika



(f) rocková skupina

Obrázek 5.3: Spektrogramy testovaných vzorků audiosignálu. Zobrazeny jsou koeficienty v Gaborově framu generovaném Hannovým oknem délky $w = 1024$ vzorků s posunem $a = 256$ vzorků a $M = 1024$ kmitočtovými pásmy.

do které by zasahovalo okno framu, které se současně (částečně) překrývá s částí signálu s uměle sníženou amplitudou. V úvahu zde bereme maximální délku díry odvozenou výše (1790 vzorků, což odpovídá přibližně 40 ms) a maximální délku okna mezi zvolenými framy dle tabulky 5.1 (2048 vzorků, což odpovídá 46 ms).

Všechny takto vytvořené poškozené signály následně rekonstruujeme dle postupu uvedeného v algoritmu 4, přičemž pro reprezentaci signálu využíváme některý z těsných Gaborových framů uvedených v tabulce 5.1.

Poznámka. Držíme se zde konvenčního značení, kdy w je délka nosiče okna, a je posun mezi okny, M je počet kmitočtových pásem a A a B jsou optimální horní a dolní mez framu (dle definice 10), které se dle předpokladu těsnosti framu rovnají. Značení typu oken je převzato z [16].

Tabulka 5.1: Parametry těsných Gaborových framů použitých v experimentu

okno	w	a	M	$A = B$
Hann	512	128	512	4
Hann	1024	256	1024	4
Hann	2048	512	2048	4
sine	1024	512	1024	2
sine	1024	128	1024	8
sqrttria	1024	512	1024	2

Rekonstrukci provádíme vždy postupně s využitím všech možností generování váhovací funkce pro kompenzaci poklesu energie (dle části 4.2.6), včetně přístupu, kdy kompenzaci neuvažujeme, tedy s prahem τ konstantním.

Díky tomu, že pracujeme s primárně nepoškozenými signály, můžeme hodnotit kvalitu rekonstrukce porovnáváním s nepoškozeným signálem. Měřítkem je pro nás SNR dle vzorce (2.10), konkrétně hodnota

$$\text{SNR}(y_{\text{org}}, y_{\text{rek}}), \quad (5.11)$$

kde y_{rek} značí vektor doplněných dat a y_{org} odpovídající úsek původního (nepoškozeného) signálu. Subjektivní hodnocení poslechem a porovnání průběhů v časové oblasti a spektrogramů originálu a rekonstrukce provedeme pro časovou náročnost pouze u vybraných realizací experimentu. V těchto realizacích také případně otestujeme metodu kompenzace poklesu energie ve výsledném signálu, která byla blíže popsána v části 4.2.6 a v algoritmu 4 byla zařazena jako poslední krok.

5.2.2. Výsledky

Výsledky provedeného experimentu jsou znázorněny pomocí krabicových diagramů na obrázcích 5.4 a 5.5. Pro vykreslení byla použita funkce `boxplot` v MATLABu. Pro každý vzorek zvlášť je vykresleno rozložení SNR – do barevně odlišených skupin jsou výsledky rozděleny podle použitého framu, římskými číslicemi pak označujeme použitou kompenzační metodu dle části 4.2.6.

Význam symbolů v diagramu je následující:

- vodorovná čára označuje medián naměřených hodnot,
- box je vykreslen v rozmezí q_1 (spodní hranice) až q_3 (horní hranice),

- horní a dolní krajní bod svislé čáry odpovídají hodnotám

$$q_3 + t \cdot (q_3 - q_1), \quad q_1 - t \cdot (q_3 - q_1), \quad (5.12)$$

kde t je volený parametr (v našem případě $t = 0,5$),

- hodnoty přesahující tyto extrémy jsou vykresleny symbolem $+$.

Je-li tedy například medián (tedy druhý kvartil q_2) blíže k hornímu okraji příslušného boxu, znamená to, že hodnoty mezi kvartily q_2 a q_3 jsou rozloženy hustěji než mezi kvartily q_2 a q_1 .

Poznámka (Značení). Volbu váhování parametru DR algoritmu značíme následovně:

- I konstanta, $\tau = \frac{1}{2}$,
- II podle délky nosiče,
- III podle ℓ_1 -normy,
- IV podle ℓ_2 -normy,
- V podle energie.

Při použití váhování je vždy parametr τ roven polovině příslušné váhy. Framy označujeme systémem `okno(w,a)` dle tabulky 5.1.

Vyhodnocení diagramů

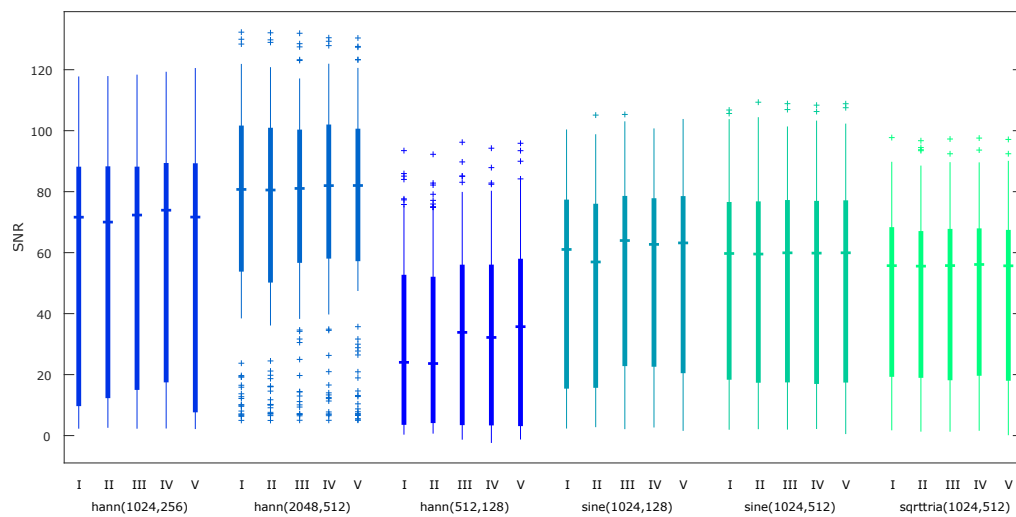
Z diagramů na obrázcích 5.4 a 5.5 vidíme, že největší vliv na kvalitu rekonstrukce pomocí navržené metody má jednoznačně řídkost signálu, což dokládá globálně klesající SNR pro jednotlivé vzorky, které byly řazeny dle řídkosti signálu. Z výsledků pro framy `hann(512,128)`, `hann(1024,256)` a `hann(2048,512)` je také zřejmý výrazný vliv délky okna. Pro soubor naměřených dat nebudeme tento trend nijak kvantifikovat, na základě diagramů však můžeme vyslovit závěr, že použitím framu s delším oknem dosáhneme obecně dle SNR lepších výsledků.

Poznámka. Zmíněné tři framy zde srovnáváme pouze s přihlédnutím k délce okna w , neboť parametr a – ačkoliv se také liší – je v těchto případech vždy dán vztahem $a = \frac{w}{4}$, není tedy nezávislý.

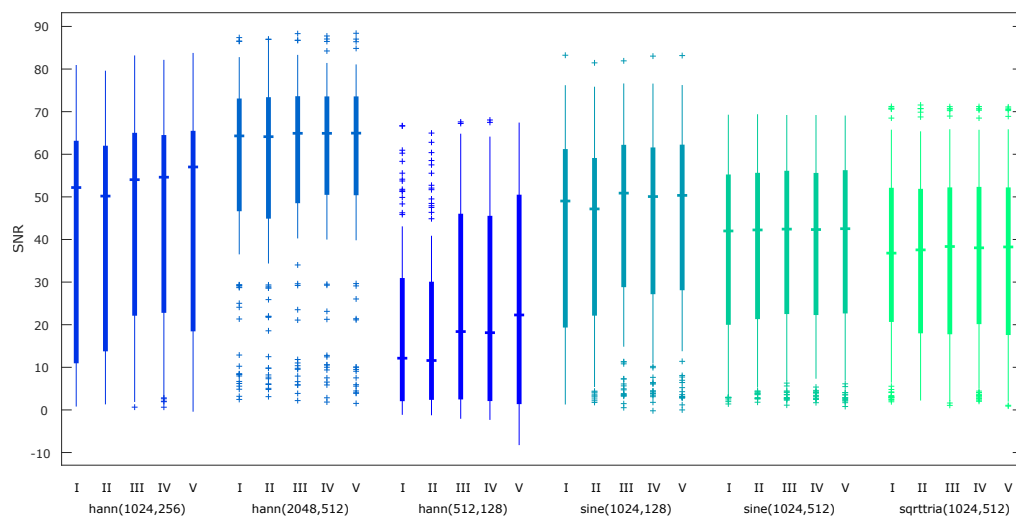
Pro soubor všech vzorků můžeme dále říci, že vliv váhování parametru τ DR algoritmu je obecně zanedbatelný při použití framů `sine(1024,512)` a `sqrtria(1024,512)` a také je nižší u méně řídkých vzorků (obrázek 5.5). Protože z diagramů není zřejmá žádná další obecná závislost SNR na volbě τ , popíšeme důsledky navržených výpočtů τ pouze zvlášť pro první tři vzorky, tedy na základě diagramů na obrázku 5.4.

Glockenspiel (zvonkohra)

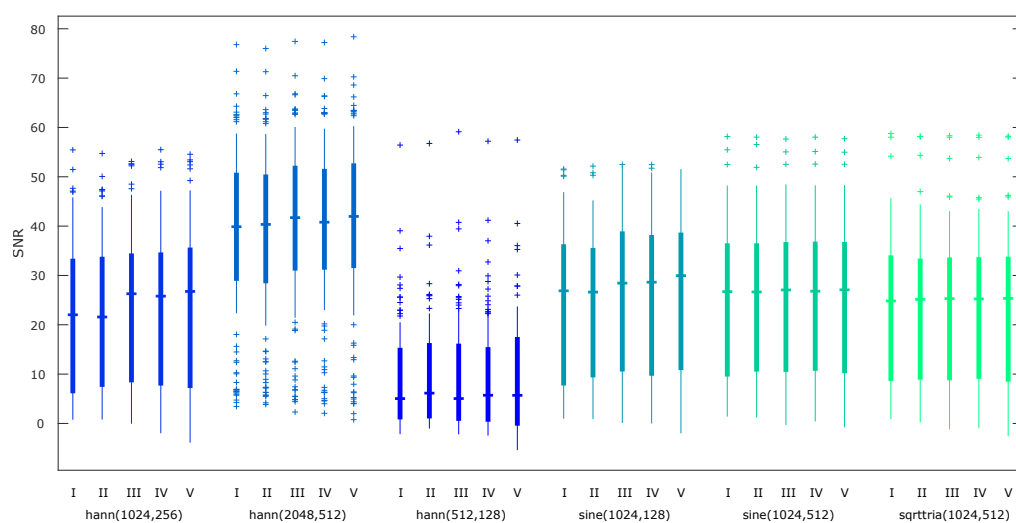
Volba τ podle ℓ_1 -normy byla nejlepší při použití framu `sine(1024,128)` – oproti τ konstantnímu se zvýšil medián SNR o přibližně 3 dB (5 %). Váhováním podle ℓ_2 -normy či podle energie jsme dosáhli výsledků srovnatelných, při váhování podle nosiče se však medián SNR snížil. Pro frame `hann(1024,256)` byla z hlediska mediánu nejlepší volba τ podle ℓ_2 -normy (zlepšení o 2,3 dB, tedy o 3 %). Největšího rozdílu ve výsledcích jsme pak dosáhli při použití framu `hann(512,128)`. Zde byla z hlediska mediánu SNR nejlepší rekonstrukce s τ podle energie (zlepšení o 11,7 dB, tedy o 49 %). Současným negativním vlivem je však zvětšení vzdálenosti mezi kvartily q_1 a q_3 , tedy větší rozptyl



(a) Glockenspiel (zvonkohra)

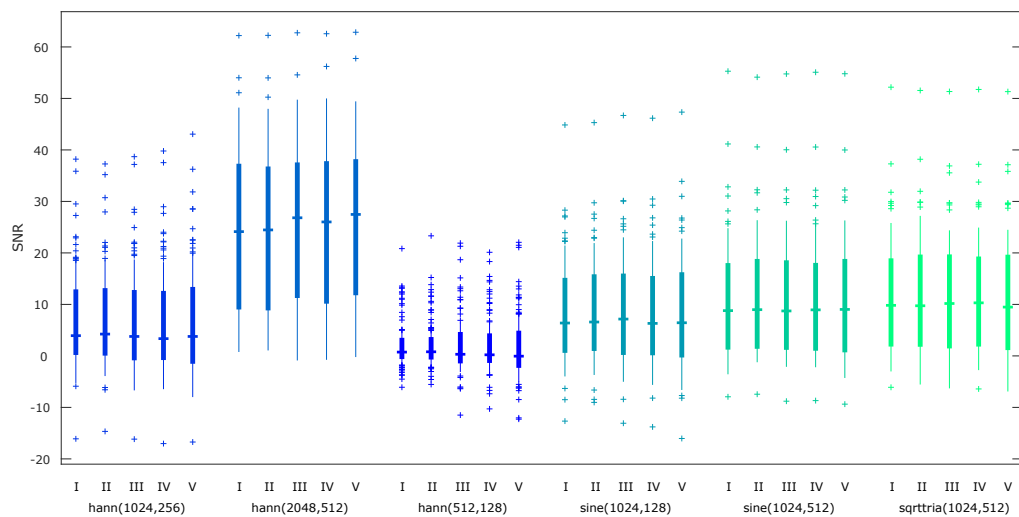


(b) housle

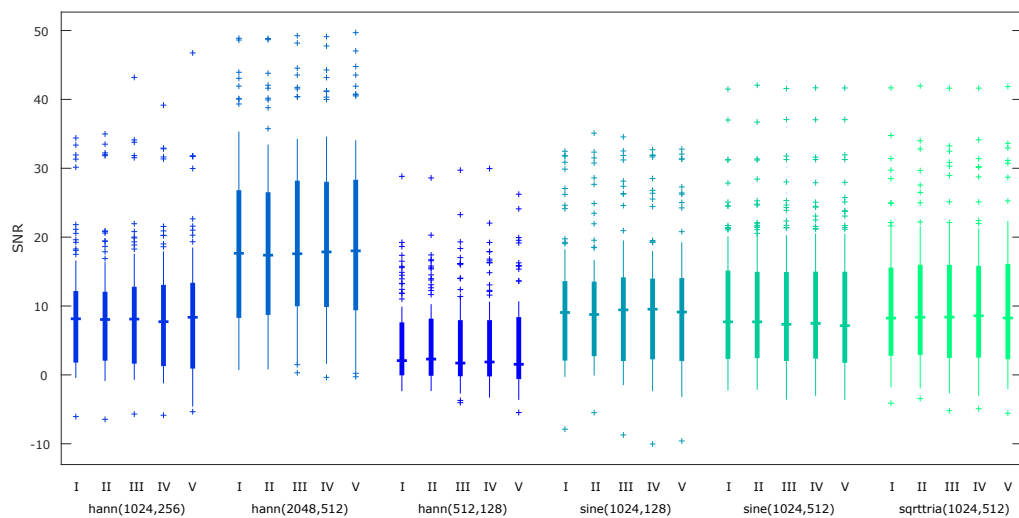


(c) housle a viola

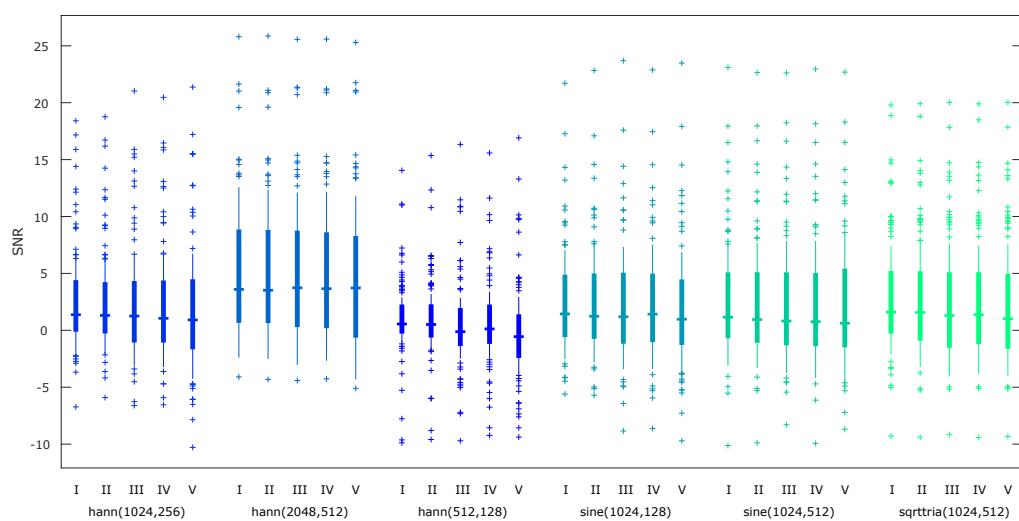
Obrázek 5.4: Krabicové diagramy – 1. část.



(a) hudecká muzika (dvoje housle, viola, kontrabas)



(b) cimbálová muzika



(c) rocková skupina

Obrázek 5.5: Krabicové diagramy – 2. část.

výsledků. Užší interval $\langle q_1, q_3 \rangle$ (tedy menší box v diagramu, například výsledky při použití framu `hann(2048,512)` ve srovnání s použitím framu `hann(1024,256)`) poukazuje na vyšší předvídatelnost výsledku metody, tedy vhodnost jejího použití pro rekonstrukci vzorků s chybějícím úsekem netestovaných parametrů.

Housle

Z hlediska mediánu se oproti konstantnímu τ nejvíce zlepšil výsledek při váhování podle energie, konkrétně u framu `hann(512,128)` to bylo o přibližně 10,2 dB (84 %). Znatelně vyšší zde byl také horní kvartil. Pozitivní vliv na rozdělení výsledků pak měla tato metoda i s použitím framu `hann(1024,256)`, kde kromě zvýšení mediánu i maxima došlo k mírnému posunu mediánu blíže k hornímu kvartilu, tzn. hodnoty byly hustěji rozloženy v oblasti s vyšším SNR.

Srovnáme-li pouze hodnoty dolního kvartilu, pak největšího zlepšení dosáhla rekonstrukce s τ podle ℓ_2 -normy (`hann(1024,256)` a `hann(2048,512)`) či podle ℓ_1 -normy (`sine(1024,128)`).

Housle a viola

U tohoto vzorku je zřetelný trend, kdy pro volby váhování parametru τ s vyšším řádem normy klesá minimum SNR, tedy váhování u příliš dlouhých děr výsledek ještě zhoršuje oproti variantě s konstantním τ . Z hlediska mediánu je u framů `hann(1024,256)`, `hann(2048,512)` a `sine(1024,128)` nejlepší rekonstrukce s váhováním podle energie – zlepšení oproti variantě s konstantním τ jsou po řadě 4,8 dB (22 %), 2,1 dB (5 %) a 3,1 dB (11 %). Srovnatelné jsou i výsledky s váhováním podle ℓ_1 - či ℓ_2 -normy.

Subjektivní hodnocení

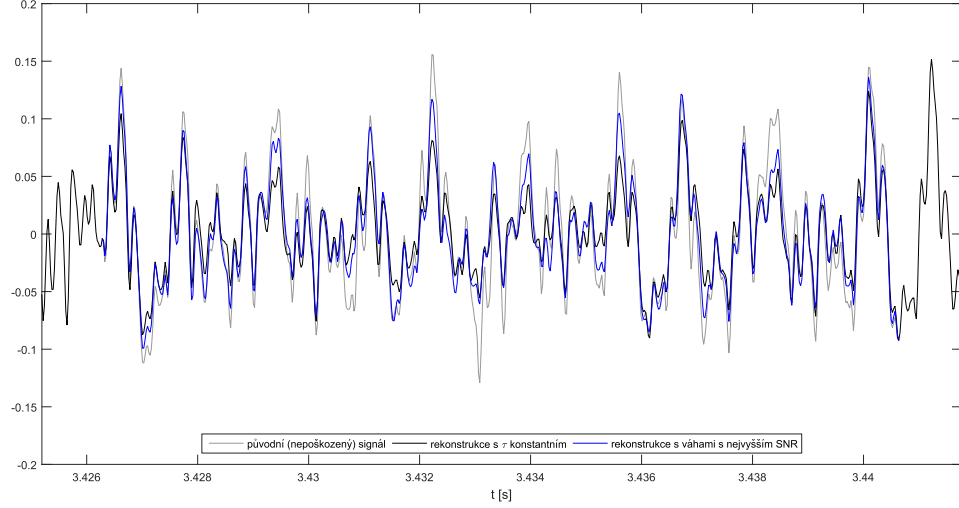
V této části vybrané výsledky experimentu popsaného výše zhodnotíme subjektivně – konkrétně poslechem originálního (nepoškozeného) a rekonstruovaného signálu a vykreslením průběhů originálu a rekonstrukce jak v časové oblasti, tak pomocí spektrogramu. Vybrané vzorky a jejich parametry jsou uvedeny v tabulce 5.2.

Tabulka 5.2: Vzorky vybrané k subjektivnímu hodnocení

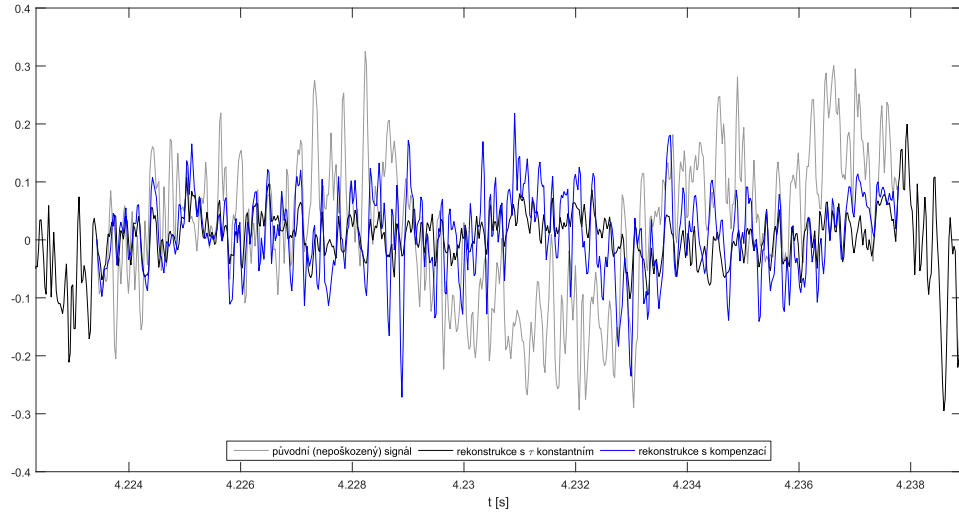
#	signál	délka díry		frame	SNR				
		vzorky	ms		I	II	III	IV	V
1	(b)	632	14,3	<code>hann(2048,512)</code>	43,9	41,0	46,8	48,4	46,6
2	(c)	632	14,3	<code>hann(2048,512)</code>	15,6	14,7	21,5	19,7	22,0
3	(d)	632	14,3	<code>hann(2048,512)</code>	7,1	7,7	13,3	10,6	12,5
4	(e)	632	14,3	<code>hann(2048,512)</code>	13,3	12,4	10,9	12,6	11,5
5	(f)	632	14,3	<code>hann(2048,512)</code>	−0,9	−0,9	−0,6	−0,8	−1,4
6	(c)	276	6,3	<code>hann(1024,256)</code>	19,1	19,5	19,0	18,4	29,5
7	(b)	1080	24,5	<code>sine(1024,128)</code>	5,4	7,9	7,3	7,6	6,4

První vzorek je ukázkou velice řídkého a stacionárního signálu. Délka díry je navíc dostatečně krátká vzhledem k délce nosiče okna framu, rekonstrukce – s parametrem konstantním i váhovaným podle ℓ_2 -normy – je od původního (nepoškozeného) signálu nerozeznatelná.

Druhý signál je již znatelně méně řídký, což je zřejmé z příslušných spektrogramů na obrázku 5.3. Pro stejnou délku úseku chybějících vzorků proto v tomto případě rekonstrukce pomocí základního algoritmu s konstantním parametrem τ obsahuje rušivý element (slyšitelné prasknutí) v místě řešené díry. Rekonstrukce s τ váhovaným podle energie je však od originálního signálu nerozpoznatelná, přestože příslušné průběhy v časové oblasti (obrázek 5.6a) se příliš neliší.



(a) signál číslo 2



(b) signál číslo 5

Obrázek 5.6: Vykreslení rekonstrukce vybraných signálů. V prvním případě byla rekonstrukce s kompenzací provedena s parametrem τ váhovaným podle energie. V druhém případě byl parametr τ váhovaný podle ℓ_1 -normy a závěrečná kompenzace v časové oblasti byla provedena pomocí funkce kosinus modulované tak, že v maximu uprostřed díry je výchylka rekonstruovaného signálu zvětšena 1,5krát.

U signálu číslo 4 jsme dle SNR dosáhli nejlepšího výsledku bez použití váhovaného parametru. Pro porovnání však byla pro subjektivní hodnocení zvolena i kompenzace s parametrem τ podle ℓ_2 -normy, která má z kompenzačních přístupů nejvyšší SNR. Poslech pak potvrdil zlepšení při použití váhovaného parametru τ a současné závěrečné kompen-

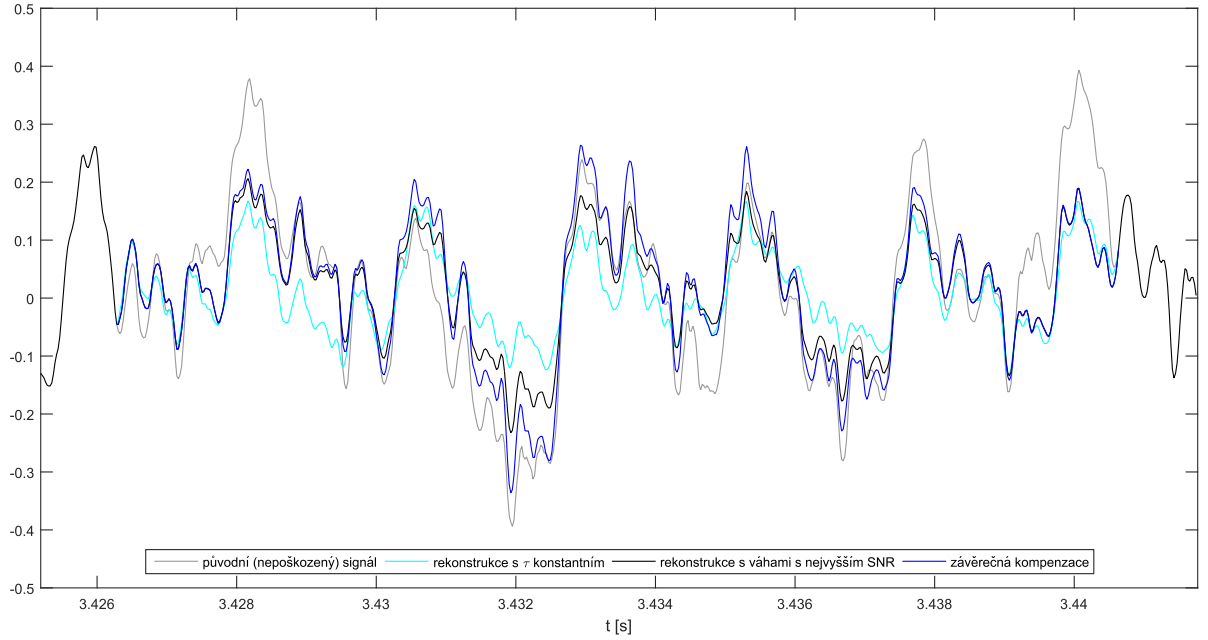
zaci oproti neváhované variantě (závěrečná kompenzace byla provedena pomocí funkce kosinus modulované tak, že v maximu uprostřed díry je výchylka rekonstruovaného signálu zvětšena 1,5krát). Stejně jako v předchozích případech je výsledná rekonstrukce (zde s použitím obou kompenzačních technik) nerozpoznatelná od neporušeného signálu.

Signál číslo 5, jehož rekonstrukce je vykreslena na obrázku 5.6b, je nejméně řídkým z vybraných vzorků. Odpovídá tomu i nízké SNR rekonstrukce a v časové oblasti výrazně se odlišující průběhy. Přesto nebyla při poslechu signálu rekonstruovaném s využitím kompenzačních metod zaplněná díra rozeznána – v tomto případě však zřejmě přispěly výrazné neharmonické složky signálu (například zvuk činelů), které rušivý pozůstatek díry skryly.

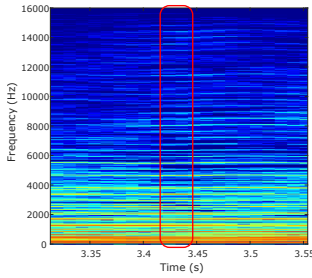
Signál číslo 6 byl vybrán pro znatelný rozdíl (dle SNR) mezi základním algoritmem a rekonstrukcí s kompenzací podle energie atomů. Tento případ je však ukázkou, kdy rozdíl v SNR nekorresponduje úplně se sluchovým vjemem, neboť oba rekonstruované vzorky byly od originálního signálu nerozpoznatelné.

Detailně nyní rozebereme zbývající dva signály. Na obrázku 5.7 je znázorněna rekonstrukce třetího signálu. V časové oblasti jsou zobrazeny postupně tři varianty rekonstrukce. Viditelné jsou zde úseky, kde rekonstruovaný signál se závěrečnou kompenzací amplitudou převyšuje signál nepoškozený. Zejména u delších úseků chybějících vzorků je tento jev problematický, neboť poukazuje na špatnou volbu funkce pro závěrečnou kompenzaci, kterou však nelze předem spolehlivě odhadnout, jak již bylo uvedeno v části 4.2.6. Poslech i v tomto případě potvrdil zlepšení při použití váhovaného parametru τ a následné zlepšení po závěrečné kompenzaci. Původní díra je zde však stále sluchem rozpoznatelná.

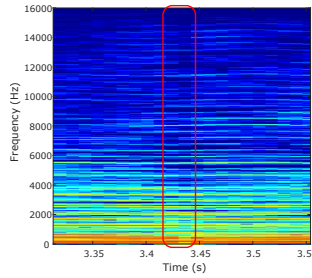
Na obrázku 5.8 vidíme průběhy signálu číslo 7. Zásadním výsledkem je zde porovnání rekonstrukce s využitím základního algoritmu a rekonstrukce váhované. Vidíme zde totiž, že s použitím kompenzační metody jsme rekonstruovali i vzorky, které by s použitím základního algoritmu zůstaly nulové. Spektrogram rekonstrukce bez závěrečné kompenzace zde není pro přehlednost vykreslen. Poslech původního vzorku a rekonstrukcí ukázal, že přestože se v časové oblasti původní signál od signálu s aplikovanými kompenzačními postupy znatelně liší, vliv kompenzace je zřetelný – rušivý výpadek signálu, který je slyšitelný ve vzorku rekonstruovaném bez použití kompenzace, je po rekonstrukci s využitím navržené kompenzační metody téměř nerozpoznatelný.



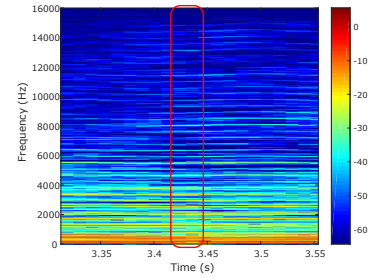
(a) průběh v časové oblasti



(b) originální signál

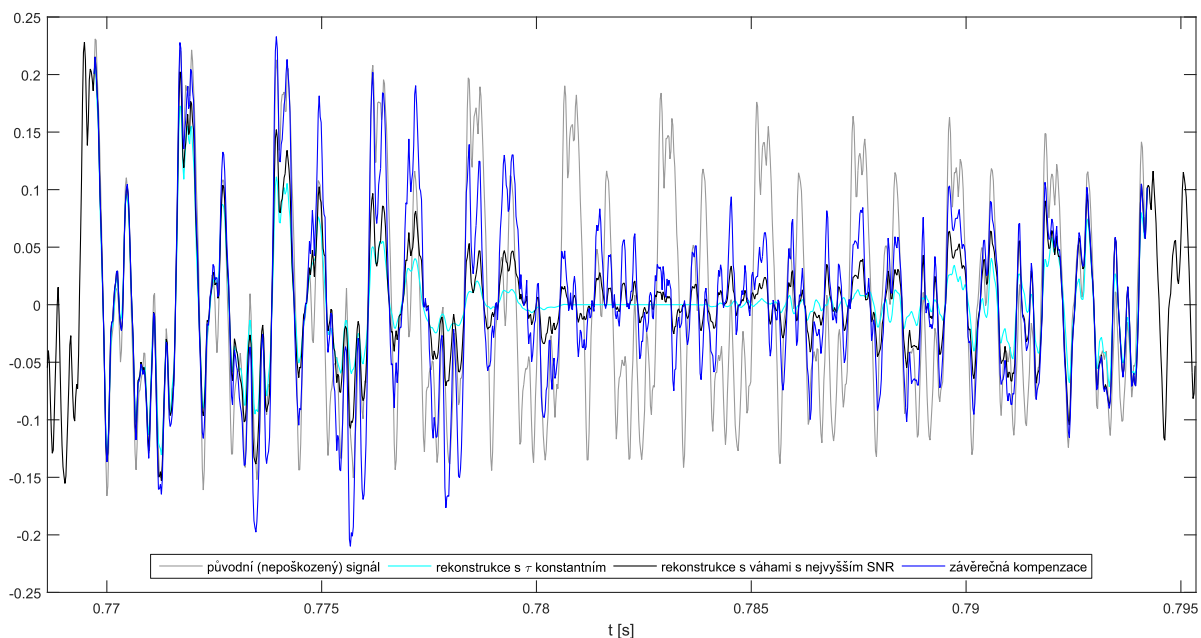


(c) rekonstrukce

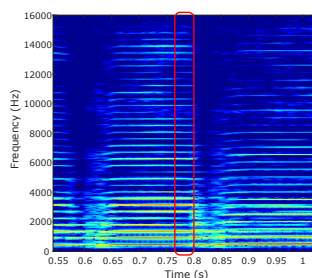


(d) rekonstrukce s kompenzací

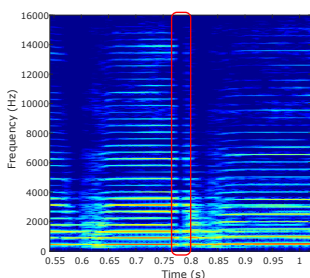
Obrázek 5.7: Vykreslení rekonstrukce signálu číslo 3. Ve všech spektrogramech je červeně znázorněn rekonstruovaný úsek chybějících vzorků. Rekonstrukce je v tomto případě s parametrem τ neváhováním, rekonstrukce s kompenzací byla provedena s parametrem τ váhováním podle ℓ_1 -normy a se závěrečnou kompenzací pomocí funkce kosinus modulované tak, že v maximu uprostřed díry je výchylka rekonstruovaného signálu zvětšena 1,5krát.



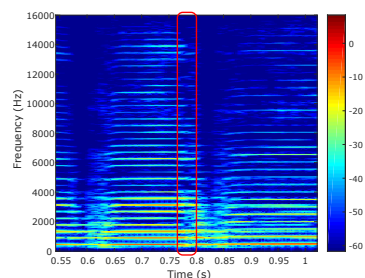
(a) průběh v časové oblasti



(b) originální signál



(c) rekonstrukce



(d) rekonstrukce s kompenzací

Obrázek 5.8: Vykreslení rekonstrukce signálu číslo 7. Ve všech spektrogramech je červeně znázorněn úsek, ve kterém se nachází díra (šířka červeného obdélníku přímo neodpovídá délce díry). Rekonstrukce je v tomto případě s parametrem τ neváhovaným, rekonstrukce s kompenzací byla provedena s parametrem τ váhovaným podle nosiče a se závěrečnou kompenzací pomocí funkce kosinus modulované tak, že v maximu uprostřed díry je rekonstruovaný signál násoben třemi.

6. Závěr

V této bakalářské práci jsme formulovali problém doplnění chybějícího úseku audiosignálu jako optimalizační úlohu. Využili jsme diskrétní Gaborovy transformace a vlastnosti, že takto získaná reprezentace audiosignálu je řídká. Řešení optimalizační úlohy bylo navrženo s využitím Douglas-Rachfordova optimalizačního algoritmu a implementováno v MATLABu. V prvních experimentech se pak objevil problém poklesu energie v rekonstruovaném signálu v místě doplněného úseku, což bylo motivací pro návrh úpravy algoritmu pro kompenzaci tohoto poklesu energie. Navrhli jsme tedy dvě základní kompenzační techniky – úpravu rekonstruovaného signálu vynásobením po bodech vhodnou funkcí a alternativně modifikaci optimalizačního algoritmu tak, aby nedocházelo k příliš velkému snižování složek reprezentace, ze kterých je následně syntetizován doplněný úsek signálu.

Následně byl navržen experiment s cílem ověřit závislost výsledků na řídkosti rekonstruovaného signálu, délce chybějícího úseku a použitém Gaborově framu. Vzhledem k tomu, že řídkost byla základním předpokladem pro navržený algoritmus, očekávali jsme, že pro méně řídké signály bude rekonstrukce méně přesná. Vzhledem k části 4.2.5 o maximální řešitelné délce díry jsme také očekávali, že lepších výsledků dosáhneme s použitím framu s delším nosičem okna. U kompenzačních postupů nebyly žádné předpoklady výsledků formulovány, neboť se jedná o originální postup. Experiment tedy měl za cíl rozhodnout o vhodnosti jednotlivých navržených přístupů ke kompenzaci.

Výsledky v části 5.2.2 jednoznačně potvrzují domněnky o vlivu řídkosti signálu, délce nosiče okna a délce doplňovaného úseku. Objektivní hodnocení na základě SNR také ukázalo zlepšení výsledků s použitím kompenzační metody v průběhu algoritmu, ovšem obecně nelze říci, která z konkrétních navržených variant je nejvhodnější. Nejzásadnější výsledek ilustruje poslední experiment ze subjektivního hodnocení v závěru páté kapitoly, neboť ukázal, že s pomocí navržené kompenzační metody můžeme rekonstruovat delší díru v signálu (oproti použití základní podoby navrženého algoritmu), aniž by se po rekonstrukci vyskytl úsek pouze nulových vzorků.

7. Seznam použitých zdrojů

Literatura

- [1] ADLER, Amir, Valentin EMIYA, Maria G. JAFARI, Michael ELAD, Rémi GRIBONVAL a Mark D. PLUMBLEY. A constrained matching pursuit approach to audio deconvolution. In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. Praha: IEEE, 2011, s. 1-5. DOI: 10.1109/ICASSP.2011.5946407. ISBN 978-1-4577-0539-7. ISSN 2379-190X.
- [2] ADLER, Amir, Valentin EMIYA, Maria G. JAFARI, Michael ELAD, Rémi GRIBONVAL a Mark D. PLUMBLEY. Audio Inpainting. *IEEE Transactions on Audio, Speech, and Language Processing* [online]. 2012, **20**(3), 922-932 [cit. 2017-04-04]. DOI: 10.1109/TASL.2011.2168211. ISSN 15587916. Dostupné z: <http://ieeexplore.ieee.org/document/6020748/>
- [3] BAHAT, Yuval, Yoav Y. SCHECHNER a Michael ELAD. Self-content-based audio inpainting. *Signal Processing* [online]. 2015, (111), 61-72 [cit. 2017-04-03]. ISSN 0165-1684. Dostupné z: <http://www.sciencedirect.com/science/article/pii/S0165168414005623>
- [4] BARATA, J. C. A. a M. S. HUSSEIN. The Moore-Penrose Pseudoinverse. A Tutorial Review of the Theory. *Brazilian Journal of Physics* [online]. Springer, 2012, **42**, 6 [cit. 2017-03-08]. DOI: 10.1007/s13538-011-0052-z. Dostupné z: <https://arxiv.org/pdf/1110.6882.pdf>
- [5] CONDAT, Laurent. A Generic Proximal Algorithm for Convex Optimization — Application to Total Variation Minimization. *IEEE Signal Processing Letters*. 2014, **21**(8), 985-989. DOI: 10.1109/LSP.2014.2322123. ISSN 1070-9908.
- [6] Proximal Splitting Methods in Signal Processing. COMBETTES, Patrick L. a Jean-Christophe PESQUET. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* [online]. New York: Springer, 2011, s. 185–212 [cit. 2017-02-04]. ISBN 978-1-4419-9569-8. Dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.298.9999>
- [7] HARRIS, F.J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE* [online]. 1978, **66**(1), 51-83 [cit. 2017-03-26]. DOI: 10.1109/PROC.1978.10837. ISSN 00189219. Dostupné z: <http://ieeexplore.ieee.org/document/1455106/>
- [8] JANSSEN, A., R. VELDHUIS a L. VRIES. Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing* [online]. 1986, **34**(2), 317-330 [cit. 2017-04-01]. DOI: 10.1109/TASSP.1986.1164824. ISSN 00963518. Dostupné z: <http://ieeexplore.ieee.org/document/1164824/>
- [9] KOVACEVIC, Jelena a Amina CHEBIRA. Life Beyond Bases: The Advent of Frames (Part I). *IEEE Signal Processing Magazine* [online]. 2007, **24**(4), 86-104

- [cit. 2017-04-01]. DOI: 10.1109/MSP.2007.4286567. ISSN 10535888. Dostupné z: <http://ieeexplore.ieee.org/document/4286567/>
- [10] PERRAUDIN, Nathanael, Nicki HOLIGHAUS, Piotr MAJDAK a Peter BALAZS. Audio inpainting with similarity graphs. In: *ArXiv.org* [online]. 2016 [cit. 2017-04-03]. Dostupné z: <https://arxiv.org/abs/1607.06667>
- [11] PFANDER, Götz E. Gabor Frames in Finite Dimensions. *Finite Frames* [online]. Boston: Birkhäuser Boston, 2013, s. 193 [cit. 2017-03-17]. DOI: 10.1007/978-0-8176-8373-3_6. ISBN 9780817683726. Dostupné z: http://link.springer.com/10.1007/978-0-8176-8373-3_6
- [12] PRŮŠA, Zdeněk, Peter L. SØNDERGAARD, Nicki HOLIGHAUS, Christoph WIESMEYR a Peter BALAZS. The Large Time-Frequency Analysis Toolbox 2.0. In: *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers* [online]. Cham: Springer International Publishing, 2014, s. 419-442 [cit. 2017-04-20]. DOI: 10.1007/978-3-319-12976-1_25. ISBN 978-3-319-12976-1. Dostupné z: http://link.springer.com/10.1007/978-3-319-12976-1_25
- [13] RAJMIC, Pavel. *Řídké a nízkohodnostní reprezentace signálů s aplikacemi*. Brno, 2014, 154 s. Habilitační práce. Vysoké učení technické v Brně.
- [14] RAJMIC, Pavel, Hana BARTLOVÁ, Zdeněk PRŮŠA a Nicki HOLIGHAUS. Acceleration of audio inpainting by support restriction. In: *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. Brno: IEEE, 2015, s. 325-329. DOI: 10.1109/ICUMT.2015.7382451. ISBN 978-1-4673-9283-9. ISSN 2157-0221. Dostupné také z: <http://ieeexplore.ieee.org/document/7382451>
- [15] RAJMIC, Pavel. Projection onto set of linear equations defined by reduced dictionary. In: *Ústav telekomunikací* [online]. Brno, 2017 [cit. 2017-03-24]. Dostupné z: http://www.utko.feec.vutbr.cz/~rajmic/papers/projection_reduced_dictionary.pdf
- [16] SØNDERGAARD, Peter L. et al. LTFAT Reference manual. In: *LTFAT: The Large Time-Frequency Analysis Toolbox* [online]. [cit. 2017-03-24]. Dostupné z: <http://ltfat.sourceforge.net/doc/ltfat.pdf>
- [17] VELDHUIS, R., A. JANSSEN a L. VRIES. Adaptive restoration of unknown samples in certain time-discrete signals. In: *ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing* [online]. Institute of Electrical and Electronics Engineers, 1985, s. 1013-1016 [cit. 2017-04-02]. DOI: 10.1109/ICASSP.1985.1168128. Dostupné z: <http://ieeexplore.ieee.org/document/1168128/>
- [18] Proofs involving the Moore–Penrose pseudoinverse. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-03-17]. Dostupné z: https://en.wikipedia.org/wiki/Proofs_involving_the_Moore-Penrose_pseudoinverse

- [19] Speech recognition: Hidden Markov models. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-04-03]. Dostupné z: https://en.wikipedia.org/wiki/Speech_recognition#Hidden_Markov_models

Použitý software

Pro implementaci algoritmu pro doplňování chybějících úseků audiosignálu byl použit následující software:

- MATLAB, verze R2015a
- toolbox LTFAT, verze 2.2.0, dostupné online: <http://ltfat.sourceforge.net/>
- funkce `min_sig_supp` [14] pro MATLAB pro zkrácení délky zpracovaného signálu, dostupné online: http://www.utko.feec.vutbr.cz/~rajmic/software/accelerate_audio_inpaint.zip

8. Seznam použitých zkratek

\mathbb{N}	množina přirozených čísel (v tomto textu včetně 0)
\mathbb{R}	množina reálných čísel
\mathbb{C}	množina komplexních čísel
A^+	Moore-Penroseova pseudoinverze matice A
A^*	hermitovská transpozice matice A
$\text{diag } x$	diagonální matice sestavená z prvků vektoru x
$\ v\ $	norma vektoru $v \in \mathbb{V}$
$\ x\ _p$	ℓ_p -norma vektoru $x \in \mathbb{C}^N$
$\langle u, v \rangle$	skalární součin vektorů $u, v \in \mathbb{V}$
$\text{dom } f$	definiční obor funkce f
prox_f	proximální operátor funkce f
$\log x$	dekadický logaritmus čísla x
w	délka nosiče okna framu (počet vzorků)
a	délka posunu oken framu (počet vzorků)
M	počet kmitočtových pásem framu
SNR	Signal-to-noise ratio
q_1	první kvartil (percentil 25)
q_2	druhý kvartil – medián (percentil 50)
q_3	třetí kvartil (percentil 75)

9. Seznam příloh

Vzorky

Přiloženy jsou vybrané použité audiosignály ve formátu WAV (ve všech případech se jedná o monofonní záznam s vzorkovací frekvencí 44 100 Hz a velikostí vzorku 16 bitů, tedy bitový tok je 705 kb/s). Označení originálních (nepoškozených) vzorků odpovídá obrázku 5.3, je tedy následující:

- (a) Glockenspiel (zvonkohra),
- (b) housle,
- (c) housle a viola,
- (d) hudecká muzika,
- (e) cimbálová muzika,
- (f) rocková skupina.

Poškozené vzorky, které sloužily pro subjektivní hodnocení, jsou označeny systémem `číslo_vzorek`, kde `číslo` je pořadí experimentu dle tabulky 5.2 a `vzorek` označuje signál dle značení výše. Díra v signálu je zde simulována úsekem nulových vzorků. Rekonstruované vzorky jsou pak značeny systémem `číslo_vzorek_metoda`, kde `metoda` značí základní algoritmus (`basic`), algoritmus s použitým váhováním parametru τ (`weighted`) nebo algoritmus s váhováním a úpravou rekonstruovaného signálu v časové oblasti (`full`).

Seznam přiložených vzorků

Nepoškozené vzorky:

a.wav
b.wav
c.wav
d.wav
e.wav
f.wav

Rekonstruované vzorky:

1_b.basic.wav
1_b.weighted.wav
2_c.basic.wav
2_c.weighted.wav
3_d.basic.wav
3_d.weighted.wav
3_d.full.wav

Poškozené vzorky:

1_b.wav
2_c.wav
3_d.wav
4_e.wav
5_f.wav
6_c.wav
7_b.wav

4_e.basic.wav
4_e.weighted.wav
4_e.full.wav
5_f.basic.wav
5_f.weighted.wav
5_f.full.wav
6_c.basic.wav
6_c.weighted.wav
7_b.basic.wav
7_b.weighted.wav
7_b.full.wav

Zdrojové kódy

Seznam příložených zdrojových kódů

<code>atom_weights.m</code>	funkce pro určení vah atomů vstupujících do díry dle části 4.2.6
<code>experiment.m</code>	ukázkový skript s experimentem číslo 7
<code>inpaint.m</code>	realizace algoritmu Douglas-Rachford (algoritmus 3)
<code>inpainting.m</code>	funkce pro rekonstrukci poškozeného signálu dle algoritmu 4
<code>inpainting_SNR.m</code>	modifikovaná funkce pro rekonstrukci, výstupem je pouze SNR
<code>min_sig_supp.m</code>	funkce pro zkrácení nosiče signálu na nutné minimum
<code>signal.mat</code>	soubor s uloženými daty pro <code>experiment.m</code>

Ukázkový skript `experiment.m` realizuje experiment číslo 7 dle tabulky 5.2. Výstupem je graf s průběhy v časové oblasti a příslušné spektrogramy. Odpovídající vzorky k poslechu jsou originální (nepoškozený) signál `b.wav` a varianty rekonstrukce `7_b_basic.wav`, `7_b_weighted.wav` a `7_b_full.wav`.